Université de Montréal
École des Hautes Études Commerciales
École Polytechnique

Université de Montréal
C. P. 6128, Succursale Centre-ville
Montréal (Québec) Canada H3C 3J7

Telephone : +1-514-343-7575
Telecopieur (Fax) : +1-514-343-7121
crt@crt.umontreal.ca
www.crt.umontreal.ca/CRT/

CRT-2002-05

Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

A SURVEY OF OPTIMIZATION MODELS FOR LONG-HAUL FREIGHT TRANSPORTATION

by

Teodor Gabriel Crainic

February 2002

PUBLICATION
CRT-2002-05

A SURVEY OF OPTIMIZATION MODELS FOR LONG-HAUL FREIGHT TRANSPORTATION*

by

Teodor Gabriel Crainic[1]

* This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR). Revised version of the CRT-98-67 entitled "Long-Haul Freight Transportation".

[1] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7 et Département de management et technologie, Université du Québec à Montréal, C.P. 6192, succursale Centre-ville, Montréal, Canada H3C 4R2 (theo@crt.umontreal.ca)

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2002

Abstract

We present the main freight transportation planning and management issues, briefly review the associated literature, describe a number of major developments, and identify trends and challenges. In order to keep the length of the paper within reasonable limits, we focus on long-haul, intercity, freight transportation. Optimization-based operations research methodologies are privileged. The paper starts with an overview of freight transportation systems and planning issues and continues with models which attempt to analyze multimodal, multicommodity transportation systems at the regional, national or global level. We then review location and network design formulations which are often associated with the long-term evolution of transportation systems and also appear prominently when service design issues are considered as described later on. Operational models and methods, particularly those aimed at the allocation and repositioning of resources such as empty vehicles, are then described. To conclude, we identify a number of interesting problems and challenges.

Key words : long-haul freight transportation, models and methods, planning, design, service design, resource allocation

Résumé

Nous présentons une revue des principaux problèmes, modèles, méthodes et outils associés à la planification et à l'opération des systèmes de transport de marchandises. L'article vise principalement le transport longue distance et privilégie les approches d'optimisation de recherche opérationnelle. L'article débute avec une description des systèmes de transport de marchandises et des principaux problèmes de planification qui leur sont associés. Sont ensuite discutées les méthodes dirigées vers la planification des systèmes de transport à l'échelle d'une région, d'un pays ou du globe. La revue des modèles de localisation et de conception de réseaux suit. Ces modèles sont fréquemment utilisés, tant dans l'analyse de l'évolution à long terme des systèmes de transport que pour la planification de réseaux de service. L'article continue par l'examen de modèles et de méthodes axés vers les problèmes opérationnels, principalement ceux associés à l'allocation et au repositionnement de ressources, les véhicules vides par exemple. Nous concluons en identifiant certains problèmes et défis intéressants d'un point de vue méthodologique et pratique.

Mots-clés : transport interurbain des marchandises, modèles et méthodes, planification, conception de réseaux de service, allocation de ressources

1 Introduction

Freight transportation is a vital component of the economy. It supports production, trade, and consumption activities by ensuring the efficient movement and timely availability of raw materials and finished goods. Transportation accounts for a significant part of the final cost of products and represents an important component of the national expenditures of any country (Crainic and Laporte, 1997).

The freight transportation industry must achieve high performance levels in terms of economic efficiency and quality of service. The former, because a transportation firm must make a profit while evolving in an increasingly open, competitive, and still mainly cost-driven market. The latter, because transportation services must conform to the high standards imposed by the current paradigms of production and management such as small or no inventory associated with just-in-time procurement, production and distribution, on-time personalized services, and customer-driven quality control of the entire logistics chain. For the transportation firm, these standards concern particularly total delivery time and service reliability, which are often translated into objectives such as "be there fast but within the specified limits" or "offer high quality service and consistent performance".

The political evolution of the world impacts the transportation sector as well. The emergence of free trade zones together with the opening of new markets due to political changes and the resulting globalization of the economy have tremendous consequences for the evolution of transportation systems, not all of which are yet apparent or well understood. For example, open borders generally mean that firms are no longer under obligation to maintain a major distribution center in each country. In consequence, distribution systems are reorganized around fewer but bigger warehouses and transportation services are operated over longer distances. A significant increase in road traffic is a normal consequence of this process, as may be observed in Europe.

Changes to the regulatory environment have an equally powerful impact on the operation and competitive environment of transportation firms. The deregulation drive of the 1980s has seen governments remove numerous rules and restrictions, especially with regard to the entry of new firms in the market and the fixing of tariffs and routes. This resulted in a more competitive industry and in changes to the number and characteristics of transportation firms. Different types of issues and regulations were being At the same time, a number of new policies and regulations resulting from quality-of-life concerns start to significantly impact the operations of the freight transportation-related firms. Two major examples: (i) more stringent safety regulations; (ii) policies aimed towards increasing the volume of inter (and multi) modal freight movements while decreasing the utilization of trucks. The latter result from environmental and energy efficiency concerns and are particularly important in Europe. The evolution of technology is another major factor that modifies how freight transportation is organized and operated. This is not a

new trend. Indeed, transportation has followed the industrial innovations and adapted, for example, to advances in traction technologies and fuels. What is new is that a, arguably the major technological factor inflecting the evolution of transportation has to do with information and software rather than the traditional hardware. The tremendous expansion of Internet and the electronic-society, eloquently illustrated by the growing importance of electronic market places and business-to-business exchanges, dramatically alters the interactions of carriers and shippers. Intelligent Transportation Systems, on the other hand, both offer means to efficiently operate and raise new challenges, as illustrated by the evolution towards real-time modification to planned routes to account for changes in traffic conditions or new demands. More complex planning and operating procedures are a direct result of these new policies, requirements, technologies, and challenges.

Freight transportation must adapt to and perform within these rapidly changing political, social, and economic conditions and trends. In addition, freight transportation is in itself a complex domain: many different firms, organizations, and institutions, each with their own set of objectives and means, make up the industry; infrastructure and even service modifications are capital-intensive and usually require long implementation delays; important decision processes are often strongly interrelated. It is thus a domain where accurate and efficient methods and tools are required to assist and enhance the analysis, planning, operation, and control processes.

The focus of the paper is on *long-haul (intercity) transportation*, that is, on transportation operations that are mainly concerned with the movement of goods over relatively long distances, between terminals or cities. Goods may be moved by rail, truck, ship, etc., or any combination of modes. The objective of the paper is to present the main freight transportation planning and management issues, to briefly review the associated literature, to describe a number of major developments, and to identify trends and challenges. In order to keep the length of the paper within reasonable limits, optimization-based operations research methodologies are privileged.

The paper is organized as follows. Section 2 presents an overview of freight transportation systems and planning issues. Section 3 is dedicated to models which attempt to analyze multimodal, multicommodity transportation systems at the regional, national or global level. Section 4 reviews network design formulations which are often associated with the long-term evolution of transportation infrastructures and services. These formulations also appear prominently when service design issues are considered as described in Section 5. Of the many operational issues related to the movement of freight, we focus on one of the most important in Section 6: the allocation and repositioning of resources, particularly empty vehicles. To conclude, Section 7 attempts to identify a number of interesting problems and methodological challenges. The Annex presents a brief survey of location models used for the strategic planning of freight transportation firms.

2 Freight Transportation Systems

Demand for freight transportation derives from the interplay between *producers* and *consumers* and the significant distances that usually separate them. Producers of goods require transportation services to move raw materials and intermediate products, and to distribute final goods in order to meet demands. *Carriers* supply transportation services. Railways, shipping lines, trucking companies, and intermodal container and postal services are examples of carriers. Considering the type of service they provide, ports, intermodal platforms, and other such facilities may be described as carriers as well. *Shippers*, which may be producers of goods or some intermediary firm (*brokers*), attribute demand to supply. Governments contribute the infrastructure: roads and highways, as well as a significant portion of ports, internal navigation, and rail facilities. Governments also regulate (e.g. dangerous and toxic goods transportation) and tax the industry.

When examining freight transportation, one often distinguishes between producers that own or operate their own transportation fleet (which then become carriers for their own freight), and "for hire" carriers, which perform transportation services for various shippers. From a planning and operations point of view, a more interesting and practical classification differentiates between: (1) Long-haul transportation (this paper) and vehicle routing and distribution, *VRP*, problems (Golden and Assad 1988, Ball *et al.* 1995, Dror 2000, Toth and Vigo 2002, etc.); (2) The multimodal transportation system of a region, irrespective of its dimensions (Section 3), and the transportation services of a particular carrier (Sections 5 and 6); (3) *Consolidation* transportation where one vehicle or convoy may serve to move freight for different customers with possibly different initial origins and final destinations, and *door-to-door* transportation operations *customized* for a particular customer.

Most freight transportation planning issues exhibit a *multicommodity* nature. In most cases, several distinct commodities must be moved. Even when the transportation system or study is dedicated to one commodity only, the traffic between different origin and destination points must be individually accounted for. Most of the time, both conditions must be satisfied simultaneously.

2.1 Customized transportation

Truckload trucking offers a typical example of door-to-door long distance transportation. In this mode, a vehicle – truck – is usually dedicated to each customer. When the customer calls, a truck with a driver or driving team is assigned to it. The truck is moved to the customer-designated location, and it is loaded. It then moves to the specified destination; this is the long-haul transportation operation. At destination, the truck is unloaded, and the driver calls the dispatcher to give its position and request a new

assignment. The dispatcher may indicate a new load, ask the driver to move empty to a new location where demand should appear in the near future, or have the driver wait and call later.

The truckload carrier thus evolves in a highly dynamic environment, where little is known with certainty regarding future demands, travel times, waiting delays at customer locations. precise positions of loaded and empty vehicles at later moments in time, and so on. Service is tailored for each customer and the timely assignment of vehicles to profitable demands is of the outmost importance.

The development of efficient *resource management and allocation* strategies are therefore at the heart of the management process. These strategies attempt to maximize the volume of demand satisfied (loads moved) and the associated profits, while making the best use of the available resources: drivers, tractor and trailer fleets, etc. Navigation services ensured by for-hire ships share some of these dynamic and stochastic characteristics.

2.2 Consolidation transportation

When demands of several customers are served simultaneously by using the same vehicle or convoy, one cannot tailor services individually for each customer. Carriers must establish regular services (e.g., a container ship from Seattle to Singapore) and adjust their characteristics (route, intermediary stops, frequency, vehicle and convoy type, capacity, speed, etc.) to satisfy the expectations of the largest number of customers possible. Externally, the carrier then proposes a series of *routes*, or *services*, each with its operational characteristics. Services are often grouped in a *schedule* that indicates departure and arrival times at the stops of the route. Internally, the carrier builds a series of rules and policies that affect the whole system and are often collected in an *operational plan* (also referred to as *load* or *transportation* plan). The aim is to ensure that the proposed services are performed as stated (or as closely as possible), while operating in a rational, efficient, and profitable way. The presence of terminals where cargo and vehicles are consolidated, grouped, or simply moved from one service to another characterizes this type of transportation performed by Less-Than-Truckload (LTL) motor carriers, railways, shipping lines, postal and express shipment services, etc. Freight transportation in some countries where a central authority more or less controls a large part of the transportation system also belongs to this category. We include all these systems under *service* or *consolidation* transportation.

The underlying structure of a large consolidation transportation system consists of a rather complex network of terminals connected by physical or conceptual links. Air and sea lines correspond to the latter, while road, highways, and rail tracks are typical examples of the former. The network may belong entirely or partially to the carrier.

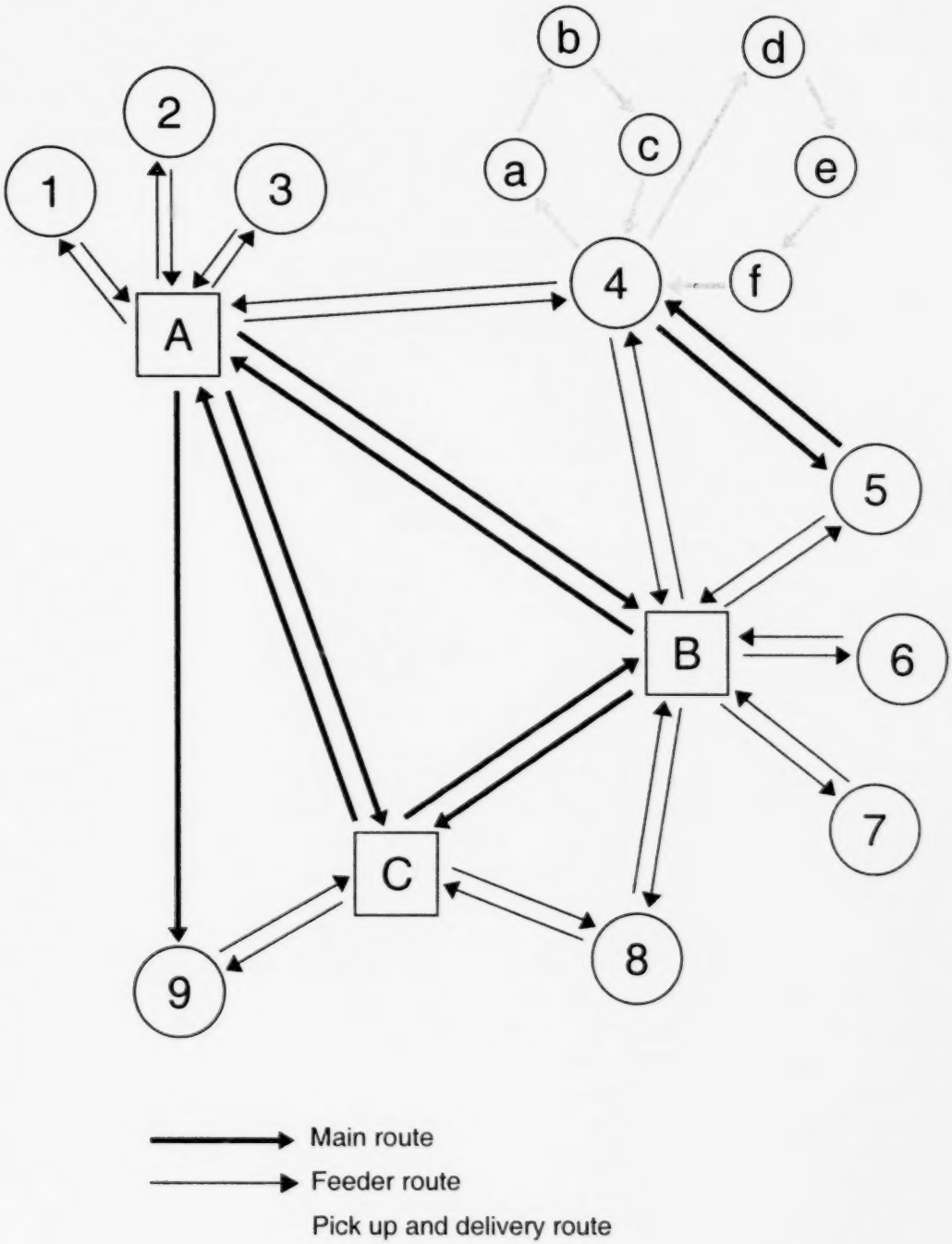

Figure 1: Network Representation of Consolidation Transportation

Rail transportation belongs traditionally to the first category, while LTL motor carriers exemplify the second: LTL carriers generally own the terminals but operate on public roads. It is noteworthy that the current policy of the European Union to separate the infrastructure ownership and the service provider (the carrier) operations is moving rail transportation in Europe towards a more LTL-like mode of operations. Some carriers prefer not to own any infrastructure, however, and only rent space as needed. Intermodal container carriers generally belong to this category, their terminal operations being often organized in ports and railway yards.

Freight demand is defined between given points of this network. Other than its specific origin, destination, and commodity-related physical characteristics, such as weight and volume, each individual shipment may present any number of particular service requirements in terms of delivery conditions, type of vehicle, and so on. A profit or cost also usually accompanies a specific demand. The consolidation carrier moves the freight by services performed by a large number of vehicles: rail cars, trailers, containers, ships, etc. Vehicles move, usually on specified routes and sometimes following a given schedule, either individually or grouped in convoys such as rail or barge trains, or multi-trailer assemblies. Convoys are formed and dismantled in terminals. Other major terminal operations include freight sorting and consolidation, its loading into or unloading from vehicles, as well as vehicle sorting, grouping, and transferring from one convoy to another. Terminals come in several designs and sizes and may be specialized in certain operations or the handling of particular products, or offer a complete set of services. In all cases, terminal operations are vital to the performance of a consolidation transportation system.

Figure 1 illustrates the network of a consolidation transportation system. Nodes A, B, and C represent major consolidation centers, also referred to as *hubs*, linked by high frequency and capacity services. Nodes 1 to 9 stand for the origin and destination terminals where freight and vehicles are consolidated at the beginning and end of the journey, and which are linked to hubs by feeder services. The figure also emphasizes the possibility for a terminal to be linked to more than one hub and illustrates the local pickup and delivery operations usually associated to terminals. Such an organization allows a much higher frequency and quality of service among hubs and a more efficient utilization of resources. The drawback is the increased delays - longer routes and more time spent in terminals - experienced by passengers or goods. This explains partly why there is hardly any "pure" hub-and-spoke systems in operation, direct transportation being organized for high demand or high priority origin-destination pairs. The links between terminals 4 and 5, and from hub A to terminal 9 in Figure 1 illustrate this option. Note that smaller firms may take advantage of consolidation systems and identify profitable niches by offering direct services to markets that large firms serve through hubs.

To further clarify these notions, consider the case of railway transportation that operates networks made up of single or double track lines that link many large and small classification yards, in which rail cars are grouped and trains are formed, pickup and

delivery stations, junction points, etc. (Assad 1980, Cordeau, Toth, and Vigo 1998). Here, everything begins when a customer issues an order for a number of empty cars or, alternatively, when freight is brought into the loading facility following a pickup operation. At the appropriate yard, rail cars are selected, inspected, and then delivered to the loading point. Once loaded, cars are moved to the origin yard (possibly the same) where they are sorted, or *classified*, and assembled into *blocks*. A block is a group of cars, with possibly different final destinations, arbitrarily considered as a single unit for handling purposes from the yard where it is made up to its destination yard where its component cars are separated. Rail companies use blocks to take advantage of some of the economies of scale related to full train loads and the handling of longer car strings in yards. The block is eventually put on a train and this signals the beginning of the journey. During the long-haul part of this journey, the train may overtake other trains or be overtaken by trains with different speeds and priorities. When the train travels on single-track lines, it may also meet trains traveling in the opposite direction. Then, the train with the lowest priority has to give way and wait on a side track for the train with the higher priority to pass by. At yards where the train stops, cars and engines are regularly inspected. Also, blocks of cars may be *transferred*, i.e., taken off one train and put on another. When a block finally arrives at destination, it is separated from the train, its cars are sorted, and those having reached their final destination are directed to the unloading station. Once empty, the cars are prepared for a new assignment, which may be either a loaded trip or an empty repositioning movement.

One source of complication in rail freight transportation is the complex nature of the main yard activities: the classification of cars and the composition of trains. The modeling of yard operations as well as that of their interactions with the rest of the system are critical components of any comprehensive rail model. It is interesting to note that, traditionally, in most rail systems cars spend most of their lifetime in yards: being loaded and unloaded, being classified, waiting for an operation to be performed or for a train to come, or simply sitting idle on a side track. Also of interest is the fact that most rail companies have separated the operations and yards dedicated to intermodal services from those used for their regular services in an attempt to cut delays, especially those associated with yard operations, and improve the quality of this time-sensitive and highly competitive service.

Similarly to rail transportation, LTL networks may encompass different types of terminals. Local traffic is picked up by "small" trucks and is delivered to *end-of-line* terminals where it is consolidated into larger shipments before long-haul movements. Symmetrically, loads from other parts of the network arrive at end-of-lines to be unloaded and moved into delivery trucks for final delivery. *Breakbulks* are terminals where traffic from many end-of-line terminals is unloaded, sorted, and consolidated for the next portion of the journey. Breakbulks are the hubs of LTL networks, as major yards are the hubs of rail transportation systems In Figure 1, nodes 1 to 9 represent end-of-lines, while nodes A, B, and C stand for breakbulk terminals.

LTL motor carrier transportation follows the same basic operational structure described for rail but on a simpler scale and with significantly more flexibility due to the fundamental difference in infrastructure: While rail transportation is "captive", trucks may use any of the existing links of the road and highway network as long as they comply with the weight regulations. Furthermore, a truck is only formed of a tractor and one or several trailers (when more than one trailer is used, these are smaller and are called "pups"). Consequently, terminal operations are generally simpler; freight is handled to consolidate outbound movements but there are no significant convoy-related operations. LTL transportation may become rather complex, however, as soon as one considers the option to use rail (the *trailer-on-flat-car* - *TOFC* - option) for long distances.

It is interesting to note that intermodal container transportation may be viewed as either door-to-door or consolidation transportation. For the customer, it is door-to-door transportation. On request, containers are delivered, loaded, moved through a series of terminals and vehicles (of which the customer has little knowledge even when the exact position of the shipment is available), and are delivered to the final destination where the goods are unloaded. For the shipping company, it is a consolidation transportation system. Containers from many customers must be moved to a port by truck, barge, or rail, or a combination thereof. There, containers are grouped and loaded on a ship that navigates a well-established route, according to a tight schedule, and delivers the containers at the destination port. From there, a land transportation system delivers the containers to the final destination by using a variety of modes and terminals. Container transportation systems that operate exclusively on land may also be encountered. In this case, rail trains and inland terminals usually play the role of ships and ports. The continuous increase in the size of container ships operated on international lines exacerbates the consolidation characteristics of intermodal container transportation systems. Indeed, the huge size of the newest generation of container ships forbids them from entering many ports and makes routes with many stops uneconomical. Consequently, long-course ships stop only at a selected number of important ports - the hubs -, while smaller vessels and land transportation modes ensure delivery of containers to the other ports and final destinations.

A similar argument may be made for express letter and small package services. For customers, it is obviously a door-to-door, high quality and reliable service. For the company, it is a consolidation transportation system that usually makes use of various air, truck, and rail services. The company implements a VRP-type of service to interact with its customers and collect and distribute letters and packages. The collection and distribution centers where mail is sorted and consolidated play a role similar to that of end-of-line terminals in LTL transportation. To reach its destination, a letter or package usually passes through at least one major hub. These terminals do for express mail services what breakbulks do for LTL motor carriers. To link its national hubs and major collection and distribution centers, the company may operate its own planes, as well as use scheduled passenger flights or train services. When distances are moderate, trucks

may be used as well.

2.3 Empty flows

A constant characteristic of any freight transportation system is the need to move empty vehicles. This follows from the imbalances that exist in trade flows and that result in discrepancies between vehicle supply and demand in various zones or terminals of the system.

To correct these differences, vehicles must be moved, *repositioned*, in order to have them available to satisfy the demand of the next period. Some repositioning decisions are straightforward. When, for example, unit trains are used to move coal or iron ore from mining fields to the port on the only rail line linking the two, cars, once unloaded, are simply formed into a return train. In most cases, however, the decision of *how many* and *where* to send vehicles appears much more complicated. The alternatives are many, due to the numerous possibilities for movement and the uncertainty of future supply and demand. The search for the most economic *empty repositioning* or *empty balancing* strategy is thus a significant problem in itself, and we will find the preoccupation with these issues in many of the problems and models addressed in the following sections.

2.4 Service schedules

Another notion often encountered in transportation planning has to do with *schedules* and *scheduled services*.

In the general sense, a schedule specifies timing information for each possible occurrence of a service during a given time period: departure time at the origin, arrival/departure time information at intermediary stops, and arrival time at the final destination. The schedule may also include indications on the *cut-off* time: the latest moment freight may be given to the carrier and still meet the scheduled departure of the service. Schedules are omnipresent in passenger transportation by air, rail, bus or ship and are strictly enforced (most of the time). The case is less clear for freight transportation. On the one hand, there are no schedules in door-to-door transportation, except for cut-off times. At the other end of the spectrum, regular navigation lines usually operate according to strict schedules (high port utilization fees constitute an important incentive to follow the schedules). Much air cargo is moved on passenger planes and therefore follows strict schedules. All-cargo air services are also usually operated according to well-established schedules.

LTL trucking follows much less stringent rules. Many carriers operate on a "go when

full" policy. Alternatively, earliest and latest departures may be planned, as well as the distribution of departures during the evening, which usually is the busiest period. The goal of this process is to offer customers late cut-off times and to ensure that trucks arrive at destination terminals within certain limits – at the opening of business in the morning, for example. The focus on increased customer service and tighter operations (including crew schedules) is increasing the utilization of scheduled services, however. Actually, schedules are build for part of the traffic only, representing the regular part of operations. Departures may then be added or cancelled to adjust for each day's particular conditions. In all cases, the dispatcher is responsible for orchestrating the operations, as well as for avoiding empty movements.

The tradition in most rail systems around the world was to follow some variant of the "go when full" rule. Even when schedules were prepared, they were mostly indicative of the ideal departure times and served as a basis for various dispatching rules for yard masters (e.g. "a train may leave one hour before planned departure if full and conditions down the line are appropriate"). The high volume of passenger trains already in the system, as well as the desire to decrease total transit time and improve connections, has pushed European rail companies toward more stringent schedules for their freight trains. Some companies operate according to fixed schedules and bookings similar to the ones used for passenger transportation. In recent years, North American companies have also migrated toward scheduled service operations (at least for part of their traffic) with various degrees of rapidity and success. The issues are different for overloaded systems, such as the Indian and Chinese railways, where the demand for passenger and freight transportation significantly exceeds the capacity of the system. In such environments, the emphasis is less on "scheduling" and more on managing the train and line operations to operate freight trains in between the passenger traffic.

2.5 Planning levels

Transportation systems thus appear as rather complex organizations that involve a great deal of human and material resources and that exhibit intricate relationships and trade-offs among the various decisions and management policies affecting their different components. It is convenient to classify these policies according to the following three *planning levels*:

1. *Strategic* (long-term) planning at the firm level typically involves the highest level of management and requires large capital investments over long-term horizons. Strategic decisions determine general development policies and broadly shape the operating strategies of the system. These include the design of the physical network and its evolution, the location of major facilities (e.g., terminals), the acquisition of major resources such as motive power units, and the definition of broad service

103-03112/

and tariff policies.

Strategic planning also takes place at the international, national and regional levels, where the transportation networks or services of several carriers are simultaneously considered. National or regional transportation departments, consultants, international shippers and forwarders, for example, engage in this type of activity. Sections 3 and 4 present models aimed at strategic issues at the system and firm levels, respectively.

2. *Tactical* (medium-term) planning aims to determine, over a medium-term horizon, an efficient allocation and utilization of resources to achieve the best possible performance of the whole system. Typical tactical decisions concern the *design of the service network* and may include issues related to the determination of the routes and types of service to operate, service schedules, vehicle and traffic routing, repositioning of the fleet for use in the next planning period. Tactical planning models are the object of Section 5.

3. *Operational* (short-term) planning is performed by local management, yard masters and dispatchers, for example, in a highly dynamic environment where the time factor plays an important role and detailed representations of vehicles, facilities and activities are essential. Important operational decisions concern: the implementation and adjustment of schedules for services, crews, and maintenance activities; the routing and dispatching of vehicles and crews; the dynamic allocation of scarce resources. Section 6 addresses operational planning issues.

This classification highlights how data flows among decision-making levels and how policy guidelines are set. The strategic level sets the general policies and guidelines for decisions taken at the tactical level, which determines goals, rules and limits for operational and real-time decisions. The data flow follows the reverse route, each level of planning supplying information essential for the decision making process at a higher level. This hierarchical relationship emphasizes the differences in scope, data, and complexity among the various planning issues, prevents the formulation of a unique model for the planning of freight transportation systems, and calls for different model formulations that address specific problems at particular levels of decision making.

3 Strategic System Analysis and Planning

The focus of the models and methods presented in this section is broad: strategic planning issues at the international, national and regional level, where the movements of several commodities through the transportation networks and services of several carriers are considered simultaneously. The main questions address the evolution of a given transportation system and its response to various modifications in its environment: changes

to existing infrastructure, construction of new facilities, evolution of the "local" or international socio-economic environment resulting in modifications to the patterns and volumes of production, consumption, and trade, variations in energy prices, changes to labor conditions, new environment-motivated policies and legislation, carrier mergers, introduction of new technologies, and so on and so forth. These questions are often part of cost-benefit analyses and comparative studies of investment alternatives – especially when the available monetary resources are limited – and are asked by regional or national planning agencies and regulatory authorities, as well as international financial institutions such as the World Bank. Private firms are also interested in these questions, particularly companies involved in the financing of transportation infrastructures, or firms that plan and operate the distribution of goods using several transportation modes.

The prediction of multicommodity freight flows over a multimodal network is an important component of transportation science and has attracted significant interest in recent years. One notes, however, that, perhaps due to the inherent difficulties and complexities of such problems, the study of freight flows at the national or regional level has not yet achieved full maturity, in contrast to passenger transportation where the prediction of car and transit flows over multimodal networks has been studied extensively and several of the research results have been transferred to practice (Florian and Hearn 1995; Cascetta 2001).

A "complete" strategic planning tool identifies and represents the fundamental components of a transportation system - demand, supply, performance measures and decision criteria - and their interactions. It yields product flow volumes and associated performance measures defined on a network representation of the transportation system. It aims to achieve a sufficiently good simulation of the global behaviour of the system to both offer a correct representation of the current situation and serve as an adequate analysis tool for planned or forecast scenarios and policies. It has to be tractable and produce results that are easily accessible. This constitutes an extremely broad scope and it is thus unrealistic to believe that a single formulation, mathematical or otherwise, or a single procedure may encompass all relevant elements, address all important issues, and fulfill all goals. Consequently, a strategic planning tool appears as a set of models and procedures. Other than data manipulation (e.g., collection, fusion, updating, validation, etc.) and result analysis (e.g., cost-benefit, environmental impacts, energy consumption policies, etc.) tools, the main components are: (i) *Supply modeling* representing the transportation modes, infrastructure, carriers, services, and lines; vehicles and convoys; terminals and inter-modal facilities; capacities and congestion; economic, service, and performance measures and criteria. (ii) *Demand modeling* that captures the product definitions, identifies producers, shippers, and intermediaries and represents production, consumption, and zone-to-zone (region-to-region) distribution volumes, as well as mode choices; Relations of demand and mode choice to the performance of economic policies and transportation system performance are also addressed here. (iii) *Assignment* of multi-product flows (from the demand model) to the multi-mode network (the supply

representation). This procedure simulates the behaviour of the transportation system and its output forms the basis for the analyses that conduct to the specification of the strategic plan. Therefore, it has to be both precise in reproducing current situation and general to produce robust analyses of future scenarios based on forecast data.

A complete survey of demand and mode choice estimation methodologies is beyond the scope of this paper. In the following, we only cite some of the most frequently used methodologies and associated references.

The modeling of demand corresponds to an image of the economic activities of a country, production, consumption, import and export of goods. For planning purposes, its output is a series of product (or commodity group) specific demand matrices indicating the volumes to be moved from one region or zone to another. It is often completed by the modeling of mode choice, which specifies for each product and origin-destination combination on what transportation infrastructure or services the demand may be moved.

A number of countries have developed *input/output* models of their economy that serve to determine the basic production and attraction of goods (Isard 1951; Casceta 2001 and references within). In order to use an input/output model, it is necessary to disaggregate the model inputs and outputs by region and then further disaggregate them by the zonal subdivision of the national planning model. This process is complex and is usually done in an analysis and computing environment which is not necessarily integrated with that used for the supply representation and the computation of flows by product. When an input/output model is not available, the initial determination of origin-destination matrices is carried out by using national statistics on production, consumption, imports and exports combined with sectorial surveys designed to complete missing or unreliable information. This process may be tedious since one has to reconcile data from several sources that may be collected by using different geographical subdivisions or inconsistent product definitions. The results of the disaggregated input/output model or the ad-hoc estimation procedures serve for the initial computation of origin-destination matrices for each product but without a subdivision by mode.

A second class of models that is well studied for the prediction of interregional commodity flows is the *spatial price equilibrium model* and its variants (Friesz, Tobin, and Harker 1983, Harker and Friesz 1986a,b, and Harker 1987; see also Florian and Hearn 1995 or Nagurney 1993). This class of models determines simultaneously the flows between producing and consuming regions, as well as the selling and buying prices that satisfy the spatial equilibrium conditions. Simply stated, a spatial equilibrium is reached provided that for all pairs of supply and demand regions with a positive commodity flow, the unit supply price plus the unit transportation cost is equal to the unit demand price; the sum is larger than this price for all pairs of regions with no exchanges. The transportation network used in these models is usually represented in a simplistic way (bipartite networks). These models rely to a large extent on the supply and demand

functions of producers and consumers, respectively, which are rarely available and quite difficult to calibrate. There are relatively few applications of this class of models for the determination of demand by product. The few applications reported in the literature deal with specific products which have a particular importance, such as crude oil, coal or milk products.

The mode choice definition may be rather general, e.g., petroleum moves by ship and pipeline or, alternatively, extremely specific indicating the particular multi-modal path for a given product, shipper, and origin-destination pair, or anywhere in between. The level of detail of modal specification needs not to be the same for all products or inter-zonal trade flows. The specification of the mode choice for a given product may be inferred from historical data and shipper surveys or it may result from a formal description and modeling effort (Winston 1983). *Random utility models*, developed and largely used for the analysis and planning of person transportation systems, have been proposed for freight transportation as well (Cascetta 2001) but their use in actual applications is scarce. The huge number of paths that have to be explicitly generated and stored, coupled to the challenge to perform this task for forecast data, may explain this phenomenon. At aggregated levels, mode choices have been specified for particularly important product flows by explicitly surveying the major logistic chains used between pairs of macro regions.

Once modal origin-destination matrices have been developed by some means, the next step is to assign them to the network (supply) model by using some route choice mechanism. The results of such an assignment model - product flows and performance measures - form part of the input to demand and cost-benefice modeling and analysis. The actual assignment mechanism may be based on further application of random utility models to the choice of pre-defined paths over a multi-modal network or on network optimization models. It is noteworthy that the attributes of pre-defined paths are determined by the state of the network at generation time and are not responsive to assignment results. Thus, for example, congestion conditions are very difficult to represent. Moreover, the utility and choice models have to be calibrated, and all paths have to be generated, for each scenario, which is quite difficult to perform when forecast data is used.

Another class, *network optimization models*, is generally considered to be more appropriate for the type of planning issues considered here. These formulations enable the prediction of multicommodity flows over a multimodal network that represents the transportation facilities at a level of detail appropriate for a nation or region but with relatively little abstraction. The demand for transportation services is exogenous and may originate from an input-output or spatial equilibrium model, if one is available, or from other sources, such as observed demand or scaling of past observed demand. The choice of mode or subsets of modes used are exogenous and intermodal shipments are permitted. Within the specified mode choice, the optimization (assignment) engine determines the best (with respect to the specified network performance measures) multi-modal paths for each product and origin-destination pair. In this sense, these models may be integrated

with econometric demand models as well. The emphasis is on the proper representation of the network and its several transportation modes, the corresponding intermodal transfer operations, the various criteria used to determine the movement of freight, the interactions and competition for limited resources captured through the representation of congestion effects, and the associated estimation of the traffic distribution over the transportation system considered to be used for comparative studies or for discrete time multi-period analyses.

Studies in the 1970's used rather simple network representations. Guélat, Florian, and Crainic (1990) and Crainic *et al.* (1990) review and discuss these efforts. Several studies also attempted to extend spatial equilibrium models to include more refined network representations and to consider congestion effects and shipper-carrier interactions. Friesz, Gottfried, and Morlok (1986) present a sequential model which uses two network representations: detailed separate networks for each carrier, and an aggregate, shipper-perceived network. On each carrier network commodities are transported at the least total cost. On the shipper-perceived network, traffic equilibrium principles are used to determine the carriers that shippers choose to move their traffic. This approach has proven quite successful in the study of logistics of products where a very limited number of shippers and carriers interact and strongly determine the behavior of the system. A typical example is the coal market between the electric utilities in the United States and their suppliers in exporting countries. Friesz and Harker (1985), Harker and Friesz (1986), Harker (1987, 1988), and Hurley and Petersen (1994) present more elaborate formulations. This line of research has not, however, yet yielded practical planning models and tools, mainly because the formulations become too large and complex when applied to realistic situations.

The modeling framework we present is based on the work of Guélat, Florian, and Crainic (1990). The formulation does not consider shippers and carriers as distinct actors in the decisions made in shipping freight. The level of aggregation appropriate for strategic planning of freight flows results in origins and destinations that correspond to relatively large geographical areas and leads to the specification of supply and demand representing, for each of the products considered, the total volumes generated by all the individual shippers. Furthermore, demand for strategic freight studies are often determined from data sources (national freight flow statistics, economic input/output models) which enable the identification of the mode used, but do not contain information on individual shippers. It is thus assumed that the shipper's behavior is reflected in the origin to destination product matrices and in the specification of the corresponding mode choice.

The modeling framework is that of a multimodal network, made up of modes, nodes, links, and intermodal transfers, on which multiple products are to be moved by specific vehicles and convoys between given origin and destination points. Here, a mode is a means of transportation having its own characteristics, such as vehicle type and capacity,

as well as specific cost measures. Depending on the scope and level of detail of the strategic study, a mode may represent a carrier or part of its network representing a particular transportation service, an aggregation of several carrier networks, or specific transportation infrastructures such as highway networks or ports.

The network consists of nodes $\mathcal{N}$, links $\mathcal{A}$, modes $\mathcal{M}$, and transfers $\mathcal{T}$ that represent all possible physical movements on the available infrastructure. To capture the modal characteristics of transportation, a link $a \in \mathcal{A}$ is defined as a triplet (i, m, j), where $i \in \mathcal{N}$ is the origin node, $j \in \mathcal{N}$ is the destination node, and $m \in \mathcal{M}$ is the mode allowed on the arc. Parallel links are used to represent situations where more than one mode is available for transporting goods between two adjacent nodes. This network representation is compact and enables easy identification of the flow of goods by mode, as well as various cost functions (e.g. operating cost, time delay, energy consumption, emissions, noise, risk, etc.) by product and mode. Furthermore, the network model resembles the physical network, since, for example, the rail and road infrastructures are physically different. Also, when there are two different types of services on a physical link, such as diesel and electric train services on rail lines, a separate link may be assigned to each service to capture the fact that they have different cost and delay functions. To model intermodal shipments, one must allow for mode transfers at certain nodes of the network and compute the associated costs and delays. Intermodal transfers t at a node of the network are modeled as link to link, hence mode to mode, allowed movements. A path in this network then consists of a sequence of directed links of a mode, a possible transfer to another mode, a sequence of directed links of the second mode, and so on. A transfer thus belongs to path if the two arcs that define it belong to the path. This representation allows for the restriction of flows of certain commodities to subsets of modes (e.g. iron ore may be shipped only by rail and ship) to capture the restrictions that occur in the operation of freight networks and transshipment facilities.

A product is any commodity (or collection of similar products) – goods or passengers – that generates a link flow. Each product $p \in \mathcal{P}$ transported over the multimodal network is shipped from certain origins $o \in \mathcal{N}$ to certain destinations $d \in \mathcal{N}$ within the network. The demand for each product for all origin-destination pairs is exogenous and is specified by a set of O-D matrices. The mode choice for each product is also exogenous and is indicated by defining for each O-D matrix a subset of modes allowed for transporting the corresponding demand. For example, one may indicate that the traffic out of certain regions must use rail, while in other regions there is a choice between rail and barges. This allows to capture the mode restrictions that occur in the operation of freight networks and transshipment facilities. Let $g^{m(p)}$ be a demand matrix associated with product $p \in \mathcal{P}$, where $m(p) \subseteq \mathcal{M}$ is the subset of modes that may be used to move this particular part of product p.

The flows of product $p \in \mathcal{P}$ on the multimodal network are the decision variables of the model. Flows on links $a \in \mathcal{A}$ are denoted by v_a^p and flows on transfers $t \in \mathcal{T}$ are

denoted by v_t^p; v stands for the vector of all product flows. Cost functions are associated with the links and transfers of the network. For product p, the respective average cost functions $s_a^p(v)$ and $s_t^p(v)$ depend on the transported volume of goods. Then, the total cost of product p on arc a is $s_a^p(v)v_a^p$, and it is $s_t^p(v)v_t^p$ on transfer t. The total cost over the multimodal network is the function F, which is to be minimized over the set of flow volumes that satisfy the flow conservation and nonnegativity constraints:

$$F = \sum_{p \in P} \left(\sum_{a \in \mathcal{A}} s_a^p(v) v_a^p + \sum_{t \in \mathcal{T}} s_t^p(v) v_t^p \right). \tag{1}$$

Let $\mathcal{L}_{od}^{m(p)}$ denote the set of paths that for product p lead from origin o to destination d using only modes in $m(p)$. The path formulation of the flow conservation equations are then:

$$\sum_{l \in \mathcal{L}_{od}^{m(p)}} h_l = g_{od}^{m(p)} \qquad o,\ d \in \mathcal{N},\ p \in \mathcal{P},\ m(p) \subseteq \mathcal{M}, \tag{2}$$

where h_l is the flow on path $l \in \mathcal{L}_{od}^{m(p)}$. These constraints specify that the total flow moved over all the paths that may be used to transport product p must be equal to the demand for that product. The nonnegativity constraints are:

$$h_l \geq 0, \qquad l \in \mathcal{L}_{od}^{m(p)},\ o,\ d \in \mathcal{N},\ p \in \mathcal{P},\ m(p) \subseteq \mathcal{M}. \tag{3}$$

The relation between arc flows and path flows is $v_a^p = \sum_{l \in \mathcal{L}^p} \delta_{al} h_l,\ a \in \mathcal{A},\ p \in \mathcal{P}$, where $\mathcal{L}^p$ is the set of all paths that may be used by product p, and $\delta_{al} = 1$ if $a \in l$ (and 0, otherwise) is the indicator function which identifies the arcs of a particular path. Similarly, the flows on transfers are $v_t^p = \sum_{l \in \mathcal{L}^p} \delta_{tl} h_l,\ t \in \mathcal{T},\ p \in \mathcal{P}$, where $\delta_{al} = 1$ if $t \in l$ (and 0, otherwise). Then, the system optimal multiproduct, multimodal assignment model consists of minimizing (1), subject to constraints (2) and (3). The optimality principle ensures that in the final flow distribution, for each product, demand matrix, and origin-destination pair, all paths with positive flows will have the same marginal cost (lower than on the other paths). The algorithm developed for this problem exploits the natural decomposition by product and results in a Gauss-Seidel-like procedure which allows the solution of large size problems in reasonable computational times (Guélat, Florian, and Crainic 1990).

This network model allows for a detailed representation of the transportation infrastructure, facilities and services, as well as the simultaneous assignment of multiple products on multiple modes. Vehicle and convoy traffic on the links (and transfers) of the network is deduced from the assigned product flows and is used to evaluate congestion conditions and to compute costs. Capacities are considered through congestion or penalty functions. Thus, the model captures the competition of products for the service capacity available, a feature of particular relevance when alternative scenarios of network capacity expansion are considered. It allows for the specification and combination of a wide variety of performance measures and assignment criteria, including user-optimum type of functions when the nature of a particular product requires it. Furthermore, the

model is sufficiently flexible to represent the transport infrastructure of one carrier only.

This model and algorithm are embedded in the *STAN* interactive-graphic system where they are complemented by a large number of tools to input, display, analyze, modify, and output data; specify the network and assignment models; analyze flows, costs, and commodity routings and paths. Matrix-based computing tools may be used to implement a whole gamut of mode choice and demand models. A network calculator can be used to combine network data to implement various performance and analysis models. A path analysis capability allows the visualisation and handling of paths used in assignment and the construction of demand and network performance models based on paths. The data required by the STAN system is organized into a strictly structured data bank. A macro language can be used to program complex operations and procedures. See Larin *et al.* (2000) for a detailed description of the STAN system, components, interfaces, and tools. The STAN system has been applied successfully in practice for scenario analysis and planning, and several agencies and organizations in a number of countries around the world use it. Crainic, Florian, and Léal (1990) present the application of this methodology to the study of freight rail transportation, while several other applications are discussed in Guélat, Florian, and Crainic (1990), Crainic *et al.* (1990), Crainic, Florian, and Larin (1994), Crainic *et al.* (1998, 2002).

4 Logistics Network Design

For freight carriers, strategic decisions determine general development policies and broadly shape the operating strategies of the system over relatively long-term horizons. Several such decisions affect the design of the physical infrastructure network: where to locate facilities such as loading and unloading terminals, consolidation centers, rail yards, or intermodal platforms; what type of equipment to install in each facility; on which lines to add capacity; what type of lines or capacity to add; what lines or facilities to abandon; and so on. These issues, which may be collectively identified as *logistics system design*, are the subject of this section.

Logistics system design issues are often addressed by evaluating alternatives using network models for the tactical (Section 5) or operational (Section 6) planning of transportation activities. When formal models are proposed, these generally appear either as *location* or *network design* formulations. An extensive literature exists on both subjects, addressing the analysis of formulations, the development of algorithms, and the performance of applications for a broad range of problems and issues. Location models are the object of the Annex, as well as of Mirchandani and Francis (1990), Daskin (1995), Drezner (1995), and Labbé, Peeters, and Thisse (1995). Labbé and Louveaux (1997) present an annotated bibliography concerning discrete location problems.

In the following, we focus on network design. We give a number of main references and present a general formulation together with a few extensions that may be used in freight transportation planning. For more details, the interested reader should consult the surveys by Magnanti and Wong (1984) and Minoux (1986), the discussions in Ahuja *et al.* (1995), Nemhauser and Wolsey (1988), and Salkin and Mathur (1989), and the annotated bibliography of Balakrishnan, Magnanti, and Mirchandani (1997). Survivability and connectivity issues are particularly important for telecommunication systems and the electronics industry, but may also appear prominently in the transportation industry when service must be ensured to certain regions or between particular zones. Grötschel, Monma, and Stoer (1995) survey the models and solution methods developed for this class of problems. An annotated bibliography may be found in Raghavan and Magnanti (1997).

4.1 Network Design

Network design models are extensively used to represent a wide range of planning and operation management issues in transportation, telecommunications, logistics, and production-distribution. These formulations play a particularly important role in decisions concerning the logistics structure, the service network (Section 5), and the operations (Section 6) of long distance freight transportation systems.

Network design models are defined in terms of a network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ represents the set of *nodes* or *vertices*. Demand for transportation exists at some of these nodes. The set of *arcs* or *links* $\mathcal{A} = \{a = (i,j) |\ i,j \in \mathcal{N},\ i \neq j\}$ includes all the possible ways to move directly (no intermediate nodes) between two nodes in $\mathcal{N}$. The set $\mathcal{P}$ includes the *products* or *commodities* that may move on the network. Let i and j be node indices and p the product index.

Other than the usual characteristics – length, capacity, and cost – *fixed costs* may be associated with some or all links of the network. This indicates that as soon as one chooses to use that particular arc, one incurs the fixed cost in excess of the utilization cost, which is in most cases related to the volume of traffic on the link. The objective of network design formulation thus is to choose links in a network, along with capacities, to enable the demand for transportation to be satisfied at the lowest possible system cost computed as the total fixed cost of the selected links plus the total variable cost of using

the network. A *fixed cost network design* formulation may then take the following form:

$$\text{Minimize} \quad \sum_{(ij)\in\mathcal{A}} f_{ij}y_{ij} + \sum_{(ij)\in\mathcal{A}} \sum_{p\in\mathcal{P}} c_{ij}^p x_{ij}^p \tag{4}$$

$$\text{subject to} \quad \sum_{j\in\mathcal{N}} x_{ij}^p - \sum_{j\in\mathcal{N}} x_{ij}^p = d_i^p \qquad i \in \mathcal{N},\ p \in \mathcal{P} \tag{5}$$

$$\sum_{p\in\mathcal{P}} x_{ij}^p \leq u_{ij}y_{ij} \qquad (i,j) \in \mathcal{A} \tag{6}$$

$$(y,x) \in \mathcal{S} \qquad (i,j) \in \mathcal{A},\ p \in \mathcal{P} \tag{7}$$

$$y \in \mathcal{Y} \qquad (i,j) \in \mathcal{A} \tag{8}$$

$$x_{ij}^p \geq 0 \qquad i,j) \in \mathcal{A},\ p \in \mathcal{P} \tag{9}$$

where,

y_{ij}: integer variables modeling discrete choice design decisions. When $\mathcal{Y} = \{0,1\}^{|\mathcal{A}|}$ in relation (8), $y_{ij} = 1$ only if link $(i,j) \in \mathcal{A}$ is *open*, selected for inclusion in the final network or for capacity expansion; $y_{ij} = 0$ otherwise, indicating that the link is *closed*. When $\mathcal{Y} = \mathbf{N}_+^{|\mathcal{A}|}$, the y_{ij} variables are not restricted to $\{0,1\}$ values and usually represent the number of facilities or units of capacity installed, or the level of service offered (see Section 5 for examples in service network design);

x_{ij}^p: continuous flow decision variables indicating the amount of flow of commodity p using link (i,j);

f_{ij}: fixed cost of opening link (i,j); when $\mathcal{Y} = \mathbf{N}_+^{|\mathcal{A}|}$, the hypothesis is that a f_{ij} cost is incurred for each unit of facility installed or service offered;

c_{ij}^p: transportation cost per unit of flow of product p on link (i,j);

u_{ij}: capacity of link (i,j);

d_i^p: demand of product p at node i.

This is the *linear cost, multicommodity, capacitated* version of the network design formulation; we identify it as *MCND*. Most applications and methodological developments target the formulations where the design variables are restricted to 0 or 1 values. A number of important applications require nonlinear formulations, however, such as the frequency service network design problems presented in Section 5. Some applications also require that flow variables be restricted to integer values, thus increasing the difficulty of these problems. However, since very few methodological developments have been dedicated to such variants of the network design model, the rest of this section focuses on formulations with $\{0,1\}$ design variables, continuous flow variables, and linear costs.

The objective function (4) of the network design formulation (4)–(9) measures the total cost of the system. An interesting point of view is to consider this objective as also

capturing the tradeoffs between the costs of offering the transportation infrastructure or services and those of operating the system to channel the flow of traffic. Equation (5) expresses the usual flow conservation and *demand satisfaction* restrictions. Several demand patterns may be defined, resulting in different models. In some cases, a product may be shipped from (one or) several origins to satisfy the demand of (one or) several destinations. These are models where the supply from several origins may be substituted to satisfy a given demand and are often used in the study of the distribution of raw materials. Variants with single product origin (or destination) may also be encountered.

Demand is defined between pairs of origin-destination points in most applications. In this case, and irrespective of the number of true commodities, a product may be associated with each origin-destination pair, by an appropriate modification of the graph that makes multiple copies of the nodes where several commodities originate or terminate their journeys. Let w^p be the total demand of product p. Then,

$$d_i^p = \begin{cases} w^p & \text{if node } i \text{ is the origin of commodity } p \\ -w^p & \text{if } i \text{ is the destination of commodity } p \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Constraint (6), often identified as a *bundle* or *forcing* constraint, states that the total flow on link (i, j) cannot exceed its capacity u_{ij} if the link is chosen in the design of the network ($y_{ij} = 1$) and must be 0 if (i, j) is not part of the selected network ($y_{ij} = 0$). When the capacity is so large that it is never binding (i.e., u_{ij} is at least the largest possible flow on the link), the demand may be normalized to 1 and u_{ij} may be set to $|\mathcal{P}|$. This simplifies the formulation and corresponds to the *uncapacitated* model. Relations (8) and (9) specify the range of admissible values for each set of decision variables.

Relation (7) captures additional constraints related to the design of the network or relationships among the flow variables. Together, they may be used to model a wide variety of practical situations, and this is what makes network design problems so interesting. For example, the set $\mathcal{S}$ may represent topological restrictions imposed on the design of the network, such as precedence constraints (e.g., choose link (i, j) only if link (p, q) is chosen) or multiple choice constraints (e.g., select at most or exactly a given number of arcs from a specified subset). An important type of additional constraint reflects the usually limited nature of available resources:

$$\sum_{(i,j)\in\mathcal{A}} f_{ij} y_{ij} \leq B \tag{11}$$

These *budget* constraints illustrate a relatively general class of restrictions imposed upon resources shared by several links. Note that, quite often, budget constraints replace the fixed cost term in the objective function (4). *Partial capacity* constraints also belong to this group:

$$x_{ij}^p \leq u_{ij}^p \qquad (i, j) \in \mathcal{A},\ p \in \mathcal{P} \tag{12}$$

They reflect restrictions imposed on the use of some facilities by individual commodities.

Such conditions may be used to model, for example, the maximum quantity of some hazardous goods moved by one train or ship.

An equivalent model is the *path*-based multicommodity capacitated network design formulation *PMCND*:

$$\text{Minimize} \quad \sum_{(ij)\in\mathcal{A}} f_{ij}y_{ij} + \sum_{p\in\mathcal{P}}\sum_{l\in\mathcal{L}} k_l^p h_l^p \tag{13}$$

$$\text{subject to} \quad \sum_{l\in\mathcal{L}^p} h_l^p = w^p \qquad p\in\mathcal{P} \tag{14}$$

$$\sum_{p\in\mathcal{P}}\sum_{l\in\mathcal{L}^p} h_l^p \delta_{ij}^{lp} \le u_{ij}y_{ij} \qquad (i,j)\in\mathcal{A} \tag{15}$$

$$y_{ij}\in\mathcal{Y} \qquad (i,j)\in\mathcal{A} \tag{16}$$

$$h_l^p \ge 0 \qquad p\in\mathcal{P},\ l\in\mathcal{L}^p \tag{17}$$

where,

$\mathcal{L}^p$: set of paths for commodity p;

h_l^p : flow of commodity p on path l;

δ_{ij}^{lp} : 1, if arc (i,j) belongs to path $l\in\mathcal{L}^p$ for product p (0, otherwise);

k_l^p : transportation cost of commodity p on path l, $k_l^p = \sum_{(ij)\in\mathcal{A}} c_{ij}^p \delta_{ij}^{lp}$;

with $x_{ij}^p = \sum_{l\in\mathcal{L}^p} h_l^p \delta_{ij}^{lp}$. Constraint (7) is usually addressed when paths are built. The same mechanisms may also handle some nonlinear route costs. Furthermore, the explicit consideration of path flows may open interesting algorithmic perspectives as illustrated by the tabu search method proposed by Crainic, Gendreau, and Farvolden (2000).

Note that for any setting of the design variables, these models yield capacitated multicommodity minimum cost network flow (*CMCNF*) problems in arc and path formulations, respectively. For uncapacitated design formulations, the subproblem obtained by fixing the design variables becomes an uncapacitated multicommodity flow problem that decomposes into $|P|$ *shortest path* problems (Pallottino and Scutellà 1998). Ahuja (1997) presents an annotated bibliography of these and other network flow problems.

Several problem classes may be derived from these general formulations by an appropriate definition of the network $\mathcal{G}$ and, eventually, of constraints in $\mathcal{S}$ (Magnanti and Wong 1984). Thus, when fixed costs are attributed to nodes, one obtains *location* formulations. Constraints that require the final design to be a Hamiltonian circuit yield the *Traveling Salesman Problem (TSP)*. Different sets of constraints on the form of the optimal network design yield the *Steiner* and the *Spanning Tree* problems. The capacitated

Vehicle Routing Problems may be viewed as a special case of the capacitated spanning tree formulation. This illustrates the richness of the network design models and explains the wide range of their applications.

4.2 General Solution Methods

Although relatively simple to state, network design formulations are generally very difficult to solve. From a theoretical point of view, most design formulations are $\mathcal{NP}$-hard. It has also been observed that for capacitated models, linear relaxations yield poor approximations of the mixed-integer polytope resulting in important optimality gaps. In particular, the interplay between link capacities and fixed costs is not adequately represented by these approaches. Moreover, the network flow subproblems are often highly degenerate, increasingly so when the number of commodities becomes larger. Additional algorithmic challenges follow from the very large scale of most applications. Important results have been obtained for some problem classes; for example, uncapacitated and tree-based formulations. However, much work is still needed for more general problem settings. In the following, we point to some of these results and research challenges. The articles mentioned at the beginning of the section and the references they contain offer a more in-depth treatment of the subject.

The previous models are mixed-integer formulations that may be approached by any of the methodologies available for this class of problems (e.g., Nemhauser and Wolsey 1988 or Salkin and Mathur 1989). A widely used methodology is to relax one or several groups of constraints in a Lagrangian fashion to obtain a simpler problem (Geoffrion 1974). A sequence of multiplier adjustments and resolutions of the relaxation subproblem yields a lower bound on the optimal value of the original formulation. As for multipliers, they may be adjusted by using a nondifferentiable optimization technique, subgradient or bundle, for example (Lemaréchal 1989). *Dual ascent* is another often-used approach to obtain this lower bound. In this case, the dual formulation of the linear relaxation of the problem is the starting point. Dual variables are then iteratively and systematically increased, while conforming to the complementary slackness conditions. An upper bound on the optimal value of the design problem is obtained as the objective value of a feasible solution heuristically derived from the solution to the relaxed problem. The lower and upper bounds are then usually integrated into an implicit enumeration scheme such as the *branch-and-bound* algorithm.

The polyhedral structure of the mixed-integer network design formulation may be studied to derive *valid inequalities* (or *cuts*) to be added to the formulation. Briefly, the objective is to construct, or approximate, the convex hull of the mixed-integer programming formulation by adding valid inequalities. If one succeeds and the convex hull is found, the original problem may be solved by linear programming methods. The *cutting plane* method is based on this idea and proceeds via a succession of resolutions of the

linear relaxation of the problem and cut generations. If the convex hull can only be approximated, the bounds may be strengthened, yielding more efficient *branch-and-bound* algorithms.

In many cases, the additional complexity introduced to account for the particularities of the application at hand and the large size of the problem instance make the exact resolution of the problem impractical. Heuristics are then used to obtain solutions of, hopefully, good quality. A number of heuristics, e.g., greedly adding or dropping arcs, aim to avoid mathematical programming techniques altogether but are not very successful for capacitated models. The relaxations and dual-ascent methods presented above are also often used as heuristics with interesting results. Modern heuristics, principally *Tabu Search* (Glover and Laguna 1997), *Simulated Annealing* (Laarhoven and Aarts 1987), and *Genetic Algorithms* (Goldberg 1989), are also increasingly being applied.

Much effort has been dedicated to uncapacitated versions of the problem and significant results have been obtained. In particular, Balakrishnan, Magnanti, and Wong (1989) present a dual-ascent procedure that very quickly achieves lower bounds within 1 to 4 percent of optimality. Used in conjunction with an add-drop heuristic, the method is able to efficiently address realistically sized instances of LTL service network design problems. The attractive performance of the dual-ascent procedure has led to the development of extensions to other design formulations and applications, as illustrated by the work of Barnhart, Jin, and Vance (2000) on railroad blocking. An exact solution method for the uncapacitated multicommodity fixed charge network design formulation has been recently proposed by Holmberg and Hellstrand (1998). The authors used a Lagrangian relaxation of the demand constraint (5) with subgradient optimization to derive lower bounds. Shortest path algorithms on networks derived from the Lagrangian relaxation solutions are used to yield feasible points. The bounds are then used in a branch-and-bound enumeration scheme and the authors discuss various branching and tree search strategies. Experiments were conducted on randomly generated problems and on a number of instances present in the literature (the largest problems solved had 1000 design arcs and 600 commodities) and showed that the branch-and-bound outperformed a state-of-the-art mixed-integer code with respect to problem size and computation time.

Significant results have also been obtained for the *Network Loading* problem. In this particular version of capacitated formulations, the objective is to install, or *load*, on each design arc a number of capacitated facilities, such as different transportation services. The total cost is made up of fixed link costs to install each facility and commodity-specific transportation costs. Total cost must be minimized and the point-to-point transportation demand must be satisfied. Two restrictions characterize this class of models and make their analysis somewhat simpler. First, one may load an integral number of l different capacitated facilities on each arc. Second, the facility capacities are modular, that is, if the capacities are $C_1 < C_2 < \cdots < C_l$, then C_{i+1} is a multiple of C_i. Originating with the work of Magnanti, Mirchandani, and Vachani (1993, 1995), many efforts have

been directed toward the polyhedral study of the problem in order to determine valid inequalities and facets to strengthen the formulation (e.g., Epstein 1998). Berger *et al.* (1998) present an efficient tabu search procedure for problems with multiple facilities where the modular restriction is relaxed and flows for each origin-destination pair must follow a single path.

Very few results have been obtained on capacitated problems defined on general networks that are more difficult to solve and pose considerable algorithmic challenges. The capability to compute efficiently good bounds on the optimal value of the design problem is a prerequisite to the development of solution methods that perform on large-scale problem instances with large numbers of commodities. Lagrangian relaxation approaches have been shown appropriate to address this issue (Gendron and Crainic 1994, 1996; Holmberg and Yuan 1996; Gendron, Crainic, and Frangioni 1998). Several Lagrangian relaxations are possible, however, and many offer the same theoretical bound, which is also the bound one obtains from the strong linear relaxation of the formulation (Gendron and Crainic, 1994). From an experimental point of view, the computing efficiency and convergence properties of the bounding procedures, as well as the quality of the solution one may actually obtain, are strongly dependent upon the choice of the nondifferentiable optimization technique used to solve the Lagrangian duals, and require careful calibration. Crainic, Frangioni, and Gendron (2001) calibrate and compare subgradient (Camerini, Fratta, and Maffioli 1978; Crowder 1976) and bundle-based methods (Lemaréchal 1989; Hiriart-Urruty and Lemaréchal 1993) for the *shortest path* and *knapsack* relaxations obtained by the dualization of constraints (6) and (5), respectively. Experiments on a large set of problem instances (largest problem had 700 design arcs and 400 commodities) were used to identify strategies for the efficient design and implementation of each method. The study showed, in particular, that bundle methods converge faster toward the optimal value of the Lagrangian dual, and that they are more robust with respect to parameter calibration.

The lower bounds reported in these studies are within 9 percent of the optimum on average. Feasible solutions were obtained by using resource-based decomposition methods but these yielded poor bounds. Tabu search meta-heuristics offer currently the best procedures for determining high quality feasible solutions. Crainic, Gendreau, and Farvolden (2000) propose a tabu search metaheuristic that identifies good solutions for the path formulation (13) to (17). The method combines simplex pivot moves and column generation in a tabu search framework where the design objective (13) is used to select the next solution from among the possible candidates. Long-term memories record for each design arc the frequency of inclusion in good solutions and guide the diversification of the search. Extensive experiments, on the same set of problems also used by Crainic, Frangioni, and Gendron (2001), have shown that the method dramatically improves the solutions found by the resource decomposition method. The utilization of the cycle-based neighbourhoods proposed by Ghamlouche, Crainic, and Gendreau (2002a) promises to improve further the performances of meta-heuristics for network design. According

to this strategy, the search proceeds in the space of the design variables by moving flow of several commodities simultaneously around suitably defined residual networks. Integrating these neighbourhoods in a tabu search-based path relinking method (Glover and Laguna 1992) constitutes the current best methodology for obtaining high quality, feasible solutions to capacitated multicommodity network design problems (Ghamlouche, Crainic, and Gendreau 2002b). The average optimality gap obtained for the same set of test problems was of the order of 2 to 3 percent, according to the problem type, with a maximum gap of the order of 10 percent. These results correspond to problems for which the optimal solutions are known. Notice that all mentioned meta-heuristics also allowed the resolution of problems too hard for the standard branch-and-bound of a state-of-the-art software in terms of CPU time or memory limitations.

Very few, if any, polyhedral results exist for the general network design formulation (4)-(9). When actually used, inequalities derived for "simpler" formulations (e.g., location models, uncapacitated network design or network loading problems) are adapted to the more general formulations. See, for example, the work of Kim *et al.* (1999), who use the *cutset* inequalities initially derived for the capacitated loading problem for the design of service networks for express package delivery firms.

These inequalities state that *the total capacity of any cut must support the total demand with endpoints on the two sides of the cut* and they are certainly valid for the general formulation. We do not know, however, if they define facets or how efficient they are. We certainly do not know how to generate these cuts efficiently. Since their number is extremely high, we have little guaranty regarding the efficiency of this procedure. The same questions are also pertinent regarding the other families of cuts proposed in the literature for formulations "similar" to network design. More work is thus required to identify valid inequalities and facets for the MCND and to develop methods to automatically and efficiently generate these new constraints (the *separation* problem). The work of Chouman, Crainic, and Gendron (2001, 2002) contributes towards feeling this gap. The authors adapt and specialize to multicommodity network design a number of important families of valid inequalities. They also introduce a new familly of valid inequalities. Extensive experimentation shows that 1) not all combinations of valid inequalities are equally effective in terms of solution quality, and 2) specialized cuts and procedures yield significant gains in solution quality and, especially, computational efficiency over state-of-the-art general purpose methods.

The situation and needs are similar concerning methods to identify the optimal solution of general MCND formulations. Holmberg and Yuan (1996, 1998) propose a branch-and-bound algorithm based on the Lagrangian relaxation of the flow constraints and subgradient optimization. The results appear promising, but not conclusive, especially when the dimensions of the network and the number of commodities increase. For larger problems, Kim, Barnhart, and Ware (1999) apply a combination of heuristics to reduce the size of the problem and branch-and-bound with column and constraint generation

(the so-called *branch-and-price-and-cut*; cuts are added to the root problem only). This constitutes a very interesting overture to a promising algorithmic avenue. See Hoffman and Padberg (1993), Desrosiers *et al.* (1995), Barnhart *et al.* (1998), and Barnhart, Hane, and Vance (2000) for examples of similar algorithmic structures aimed at various complex problems that arise in transportation science and which emphasize the challenges associated with the development of such methods for the MCND.

Parallel computation may help address realistically dimensioned problem instances in reasonable times. In the case of heuristics, parallelism may also enhance the robustness of the method and improve the quality of the solutions (Crainic 2002, Crainic and Gendreau 2002, Crainic and Toulouse 2002). Applied to branch-and-bound, parallelism may be used to solve the subproblem at each node of the tree (Gendron and Crainic, 1994b) or to explore the tree in parallel (Gendron and Crainic, 1994a). Many issues still remain to be addressed in this area however. For example, the addition of cuts often destroys the "nice" structure (network, knapsack, etc.) obtained by relaxing some constraints. The relaxation of the cut constraints could them be contemplated. The issue might become even more challenging when constraints are to be generated at nodes other than the root. It is generally believed, however, that the combination of relaxations, polyhedral results, and heuristics within a parallel computation framework constitutes a promising avenue towards a comprehensive solver for capacitated, multicommodity network design.

5 Service Network Design

Service network design is particularly relevant to firms and organizations that operate consolidation transportation systems and is typically related to the planning of operations. It is usually part of *tactical planning* activities, although often it is referred to as strategic/tactical or tactical/operational according to the planning traditions and horizons of the firm. The goal is to operate efficiently to answer demand and ensure the profitability of the firm. The "supply" side of this equation implies a system-wide, network view of operations, integrating consolidation activities in terminals, and the selection, routing, and scheduling of services. On the "demand" side, the routing of freight through the network must be planned to ensure timely and reliable delivery according to the customer specifications and the carrier's own targets.

The objectives of the process are complex as well. The customer's expectations have traditionally been expressed in terms of "getting there" at the lowest cost possible. This, combined with the usual cost consciousness of any firm, has implied that the primary objective of a freight carrier was, and still is for many carriers, to operate at the lowest possible cost. Increasingly, however, customers not only expect low tariffs, but also require a high quality service, mostly in terms of speed, flexibility, and reliability. The significant increase in the market share achieved by motor carriers, mainly at the expense

of railway transportation, is due to a large extent to this phenomenon. Consequently, one of the major objectives of tactical planning is to achieve the best tradeoff between operating costs and firm profitability, and service performance measured, in most cases, by delays incurred by freight and rolling-stock or by the respect of predefined performance targets.

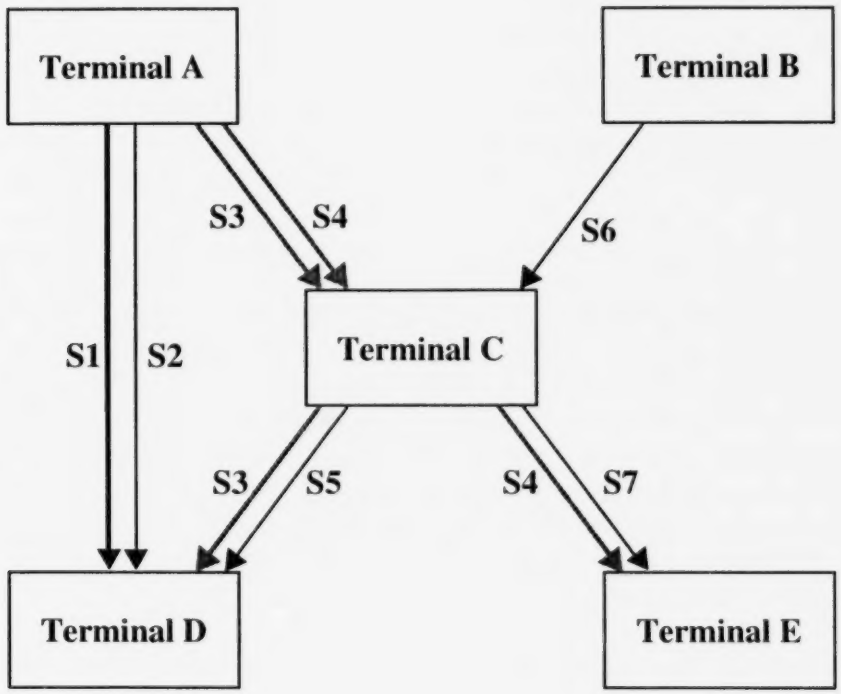

Figure 2: Service Network

To illustrate the complexity of decisions and tradeoffs characteristic of tactical planning, consider the routing of a shipment between two terminals of a consolidation transportation system operated, for example, by a railway or LTL motor carrier. Figure 2 displays a representation of such a system made up of five terminals and seven services (for simplicity, the actual service routes are not shown). A shipment that originates at terminal A with destination terminal D is sorted (classified) at A and may be routed according to a number of strategies, including:

1. Consolidate it with other shipments going directly to its destination terminal and put it on one of the available direct services, S1 or S2, of possibly different types. If the freight volume is sufficiently high and the customer contract allows it, S1

or S2 might be operated as a dedicated service, such as a full truck moving direct between two end-of-lines or an unit train.

2. Same consolidation, but move the shipment by using a service, such as S3, that stops at one or several other terminals to drop and pickup traffic.

3. Use the same consolidation policy but move the shipment by a direct service S4 to the intermediate terminal C, where it is transferred to another direct service, S5, that moves it to destination. This strategy may outperform the previous one if the service level offerred on the direct routes outweights the transfer costs; in Figure 2, it is also the only strategy available to move from terminal B to terminal D.

4. Consolidate the shipment into a load for an intermediate terminal where it will be reclassified and consolidated together with traffic originating at various other terminals into a load for its final destination. The shipment is thus moved by service S3 or S4 from A to C, consolidated together with traffic from B to D and C to D, and then moved by S3 or S5 from terminal C to destination.

Which alternative is "best"? Each has its own cost and delay measures that follow from the service characteristics of each terminal and service. Thus, for example, strategies based on reconsolidation and routing through intermediate terminals may be more efficient when direct services between the origin and destination terminals of the shipment are offered rarely due to generally low level of traffic demand. Such strategies would probably result in higher equipment utilization and lower waiting times at the original terminal; hence, in a more rapid service for the customer. The same decision would also result, however, in additional unloading, consolidation, and loading operations, creating heavier delays and higher congestion levels at intermediary terminals, as well as a decrease in the total reliability of the shipment. On the other hand, to increase the frequency of a direct service between the origin and destination terminals of a shipment would imply a faster and more reliable service for the corresponding traffic, as well as a decrease in the level of congestion at the intermediate terminals at the expense of additional resources, thus increasing the direct costs of the system. Therefore, to select the "best" solution for the customer and the company, one has to simultaneously consider the routing of all traffic, the level of service on each route, and the costs and service characteristics of each terminal. These problems and decisions have network-wide impacts and are strongly and complexly interconnected both in their economic aspects and the space-time dimensions of the associated operations. Therefore decisions should be made globally, network-wide, in an integrated manner (Crainic and Roy 1988). More formally, main decisions made at the tactical level concern the following issues:

1. *Service selection.* The routes – origin and destination terminals, physical route and intermediate stops – on which services will be offered and the characteristics of each service. *Frequency* or *scheduling* decisions are part of this process.

2. *Traffic distribution.* The *itineraries* (routes) used to move the flow of each demand: services used, terminals passed through, operations performed in these terminals, etc.

3. *Terminal policies.* General rules that specify for each terminal the consolidation activities to perform. For rail applications, these rules would specify, for example, the blocks into which cars should be classified (the *blocking* policies), as well as the trains that are to be formed and the blocks that should be put on each train (the *make up* rules). An efficient allocation of work among terminals is an important policy objective.

4. General *empty balancing* strategies, indicating how to reposition empty vehicles to meet the forecast needs of the next planning period.

Several efforts have been directed toward the formulation of tactical models. See the reviews of Assad (1980), Crainic (1988), Delorme, Roy, and Rousseau (1988), and Cordeau, Toth, and Vigo (1998). Network models, which take advantage of the structure of the system and integrate policies affecting several terminal and line operations, are the most widely developed. Simulation models have been proposed and used by transportation firms to evaluate scenarios and select policies. Network optimization formulations, on the other hand, may efficiently generate, evaluate, and select integrated network-wide operating strategies, transportation plans, and schedules. These models are discussed in this section.

Most service network design and related issues yield *fixed cost, capacitated, multicommodity network design formulations* (Section 4.1). These formulations may be static or dynamic but, up to now, have been generally deterministic. For a clearer view of tactical planning issues and service network design formulations, we distinguish between *frequency* and *dynamic* service network design models.

The former typically addresses strategic/tactical planning issues. The study and representation of interactions and tradeoffs among subsystems and decisions form a central part of this class of approaches. Typical issues addressed by such models concern questions such as: *What* type of service to offer? *How often* over the planning horizon to offer it? Which traffic *itineraries* to operate? What are the appropriate *terminal workloads* and *policies*? Frequency service network design models may be further classified according to the role service levels play in the formulations: *decision* or *output*. In a nutshell, service frequencies are explicit integer decision variables in the first class of models. Formulations that belong to the second class include "operate or not" ($\{0, 1\}$) decision variables and derive frequencies from traffic flows subject to lower bound restrictions that represent minimum service levels. The output of frequency service network design models, the *transportation* or *load plan*, is used to determine the day-to-day policies that guide the operations of the system and is also a privileged evaluation tool for "what-if" questions raised during scenario analysis in strategic planning. Dynamic formulations

are closer to the operational side of things. They usually target the planning of *schedules* and support decisions related to *if* and *when* services depart. Subsections 5.1, 5.2, and 5.3 examine models and methods that belong to each of these three classes. Section 5.4 briefly reviews the literature associated to service network design and tactical planning.

5.1 Frequency Service Network Design

The network optimization modeling framework proposed by Crainic and Rousseau (1986) constitutes a prototypical frequency service network design formulation where explicit decision variables are used to determine how often each selected service will be run during the planning period. It is a multimodal multicommodity model that integrates the service selection and traffic distribution problems with general terminal and blocking policies. Its goal is the generation of global strategies to improve the cost and service performance of the system. It is a modeling framework in the sense that while it may represent a large variety of real situations and it has to be adapted to each application. Rail applications are to be found in Crainic (1982, 1984), Crainic, Ferland, and Rousseau (1984), and Crainic and Nicolle (1986). Roy (1984) and Delorme and Roy (1989) present applications of this framework to LTL trucking. In the following, we present a simplified model in order to emphasize the main modeling issues and challenges.

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ represent the "physical network" over which the carrier operates. Vertices in $\mathcal{T} \subseteq \mathcal{N}$ correspond to nodes where the terminals selected for the particular application are situated. For simplicity, assume that all terminals can perform all operations. The *service network* specifies the transportation services that could be offered to satisfy this demand. Each *service* $s \in \mathcal{S}$ is defined by its route r_s through the physical network; origin, destination, and intermediary terminals where the service stops and work may be performed on its vehicles and cargo; capacity u^s_{ij} on each link of r_s; service class that indicates characteristics such as the mode, preferred traffic or restrictions, speed and priority of the service, etc.

Transportation demand is defined in terms of volume (e.g., number of vehicles) of a certain commodity to be moved between two terminals in $\mathcal{T}$. To simplify, we refer to product p = *(commodity type, origin, destination)* with a positive demand w^p. In the literature, one also finds the terms *market* and *traffic-class* with a similar meaning. Empty vehicles may be included as commodities to be moved between given origin-destination pairs. Traffic moves according to *itineraries*. An itinerary $l \in \mathcal{L}^p$ for product p specifies the service path used to move (part of) the corresponding demand: the origin, destination, and intermediary terminals where operations are to be performed; the sequence of services between each pair of consecutive terminals where work is performed; the commodity class that indicates characteristics such as priority, minimum service level, preferred transportation mode, etc.

Service frequencies y_s, $s \in \mathcal{S}$, define the level of service offered, i.e., how often each service is run during the planning period. To design the service network thus means to decide the frequency of each service contemplated in the planning process such that the demand is satisfied. Many itineraries may be defined for each product and more than one may be actually used, according to the level of congestion in the system and the service and cost criteria of the particular application. Flow distribution decisions are therefore represented by variables h_l^p indicating the volume of product $p \in \mathcal{P}$ moved by using its itinerary $l \in \mathcal{L}^p$. Workloads and general consolidation strategies for each terminal in the system may be derived from these decision variables.

Let $y = \{y_s\}$ and $h = \{h_l^p\}$ be the vectors containing the decision variables. The model states that the total generalized system cost must be minimized, while satisfying the demand for transportation and the service standards:

$$\text{Minimize} \quad \sum_{s \in \mathcal{S}} \Psi_s(y) + \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}^p} \Phi_l^p(y,h) + \Theta(y,h) \tag{18}$$

$$\text{subject to} \quad \sum_{l \in \mathcal{L}^p} h_l^p = w^p \qquad p \in \mathcal{P} \tag{19}$$

$$y_s \geq 0 \text{ and integer} \qquad s \in \mathcal{S} \tag{20}$$

$$h_l^p \geq 0 \qquad l \in \mathcal{L},\ p \in \mathcal{P} \tag{21}$$

where,

$\Psi_s(y)$: total cost of operating service s;

$\Phi_l^p(y,h)$: total cost of moving the freight of product p by using its itinerary l;

$\Theta(y,h)$: penalty terms capturing various relations and restrictions, such as the limited service capacity.

This model is similar to the path formulation of the capacitated network design model ((13) – (17)) introduced in Section 4.1, except that the linear cost functions of the latter have been replaced by a notation that indicates more general functional forms. The objective function defines the *total system cost* and includes the total cost of operating a service network at given frequencies, the total cost of moving freight by using the selected itineraries for each product, as well as a number of terms translating operational and service restrictions into monetary vales. $\Psi_s(y)$ and $\Phi_l^p(y,h)$ thus correspond to the *fixed* and *variable* costs, respectively, of the network formulation given the general level of service of the firm and the corresponding traffic pattern. The objective function computes a *generalized* cost, in the sense that it may include various productivity measures related to terminal and transportation operations. Other than the actual costs of performing the operations, one may thus explicitly consider the costs, delays, and other performance measures related to the quality and reliability of the service offered, to evaluate alternatives and determine the most advantageous tradeoffs.

The delays incurred by vehicles, convoys, and freight due to congestion and operational policies in terminals and on the road are generally used as a measure of service quality. Define $T_s(y)$ and $T_l^p(y,h)$ as the total durations of service s and itinerary l for product p, respectively. Equations (22) to (23) illustrate one approach to use delays to integrate service considerations into the total generalized system cost. On the one hand, unit operating costs C_s^O and C_{lp}^O are computed for each service and product itinerary, respectively. On the other hand, the corresponding total expected service, $E[T_s(y)]$, and itinerary, $E[T_l^p(y,h)]$, times are converted into measures compatible with the operating costs via unit time costs for each traffic (C_{lp}^D) and service (C_s^D) class. These costs are usually based on equipment depreciation values, product inventory costs, and time-related characteristics, such as priority or different degrees of time sensitivity for specific traffic classes.

$$\Psi_s(y) = (C_s^O + C_s^D E[T_s(y)])y_s \tag{22}$$

$$\Phi_l^p(y,h) = (C_{lp}^O + C_{lp}^D E[T_l^p(y,h)])h_l^p \tag{23}$$

Although nonlinear functions could be used, unit operation costs C_s^O and C_s^D are usually computed as the sum of the unit costs of all terminal and transportation activities described in the service routes and freight itineraries. For rail applications, these may include hauling costs for trains and cars over the lines of the network, as well as yard handling costs associated with car classification, the transferring of cars and blocks among trains, and the making-up and breaking-down of trains. Similar terms appear in LTL applications: loading, unloading, transdock, and consolidation operations at terminals, energy costs, maintenance, crews, etc.

The expected total delays $E[T_s(y)]$ and $E[T_l^p(y,h)]$ are also computed by summing up the expected delays associated with the terminal and line operations that make up the service and freight routes. No correlation is usually considered. Some durations are simply assumed proportional to the volume of vehicles or traffic handled. It is typically the case for the yard transfer delays for rail applications and intercity transportation time for LTL trucking. In many operations, however, vehicles of different services carrying freight for different products on various itineraries must use the same facilities. It is the case, for example, of most consolidation and classification operations. As a consequence, most time-related functions are built to reflect the increasingly larger delays that result when facilities of limited capacity must serve a growing volume of traffic. Such *congestion* functions are typically derived from engineering procedures and queuing models (see Section 6) and are built to represent: average delays due to rail yard operations, particularly car classification and blocking, and train make up; waiting time of trucks at LTL terminals before loading and unloading operations (rail cars and trucks at port loading/unloading facilities experience similar delays); delays incurred by trains when meeting, overtaking, or being overtaken by other trains on the lines of the network; congestion on highways in urban areas; expected departure or connection delays in rail yards, LTL terminals, and maritime ports representing the waiting time for the designated service to be available,

and so on.

Average transportation delays do not tell the whole story, however. Often, the goal is not only rapid delivery but also consistent, reliable service. The variance of the total service or itinerary time may then be used to penalize unreliable operations. Equation (24) illustrates this approach for the case when service quality targets are announced. Here, each traffic-class has a delivery objective (e.g., 24 hours) and reliability requirements (e.g., target must be achieved for 90 percent of deliveries), noted H_p and n, respectively. A penalty C^D_{lp} is then imposed when the expected itinerary time, adjusted for its standard deviation $\sigma[T^p_l(y,h)]$, does not comply with the service objective. The total itinerary cost then becomes:

$$\Phi^p_l(y,h) = C^O_{lp} h^p_l + C^D_{lp} \left(\min\{0,\ H_p - E^p_l(y,h) - n\sigma[T^p_l(y,h)]\}\right)^2 h^p_l. \tag{24}$$

Finally, equation (25) illustrates the use of penalty terms to capture various restrictions and conditions. Here, x_{sk} stands for the total volume of freight hauled by service s over its service leg k, $x_{sk} = \sum_{p\in\mathcal{P}} \sum_{l\in\mathcal{L}^p} h^p_l \delta^{lp}_{sk}$, with $\delta^{lp}_{sk} = 1$ if service leg k of service s is used by itinerary l of product p, and 0 otherwise. Thus, in this example, the service capacity restrictions are considered as utilization targets and the over-assignment of traffic is permitted at the expense of additional costs and delays. Tradeoffs between the cost of increasing the level of service and the extra costs of insufficient capacity may then be addressed while the associated mathematical programming problem is solved.

$$\Theta(y,h) = \sum_{s\in\mathcal{S}} C^P_s \sum_{(ij)\in r_s} (min\{0,\ u^s_{ij} y_s - x_{sk}\})^2 \tag{25}$$

The model has the structure of a nonlinear, mixed integer, multimodal, multicommodity network flow problem. No exact solution method has yet been proposed for this model. The original method described by Crainic and Rousseau (1986) combines a heuristic (based on finite differences in the objective function) that iteratively decreases frequencies from initial high values, with a convex network optimization procedure to distribute the freight. The latter makes use of column generation to create itineraries and descent procedures to optimize the flow distribution. The procedure appeared efficient for the rail and LTL applications considered. Crainic and Roy (1988) and Roy and Crainic (1992) also report on the utilization of this approach to perform scenario and postoptimal analyses, particularly concerning the tradeoffs between the cost of operating the system and the value of time, and the level of demand required to operate direct services over long distances.

5.2 Service Frequencies as Derived Output

The load planning model for LTL motor carriers introduced by Powell and Sheffi (1983, 1986, 1989; see also Powell, 1986a and Lamar, Sheffi, and Powell, 1990) constitutes a major example of frequency service network design formulations that yield service levels as one of their outputs. What follows is a condensed version of this model.

The model is defined on a service network $\mathcal{G} = (\mathcal{T}, \mathcal{S})$ where all nodes are terminals and links represent potential direct services between two terminals. Two types of terminals are considered: *end-of-lines*, where freight originates and terminates; and *breakbulk* consolidation terminals. Although not forbidden, direct movements between end-of-line terminals are extremely rare, especially for very large LTL carriers. Consequently, the design decisions concern only services between end-of-lines and breakbulks, and between breakbulk terminals. This has the benefit of considerably reducing the size of the problem. The main parameters and decision variables that define the model are:

C_{ij}: unit linehaul cost per trailer, loaded or empty, from terminal i to terminal j;

C_i^B: unit trailer handling cost at terminal i, if terminal i is a breakbulk (0, otherwise);

$C_i^E(\cdot)$: a function that computes the trailer handling cost at end-of-line i according to the total number of direct services operated out of i (0, if i is a breakbulk);

w_{od}: number of LTL trailers originating at terminal o and destined for terminal d;

$\mathcal{L}$: set of permissible freight routings, i.e. that respect particular constraints with respect to the association of end-of-line terminals to breakbulks (the so-called *clustering constraints*);

y_{ij}: service design decisions; $y_{ij} = 1$ if the carrier is offering direct service from terminal i to terminal j, and 0 otherwise;

x_{ij}^d: volume of LTL traffic on link (i, j) with destination terminal d; $x_{ij} = \sum_d x_{ij}^d$;

r_{ij}^d: auxiliary flow routing variable (its use simplifies the representation of the clustering constraints);

v_{ij}: flow of empty trailers moving from i to j;

x_i^B: volume of total LTL traffic handled at breakbulk i, that is, the traffic that originates at i plus the traffic that is transferred at the terminal;

M_{ij}: minimum frequency if a direct service is offered from terminal i to terminal j;

$F_{ij}(x_{ij})$: service frequency - the number of trailers dispatched over the planning period, from terminal i to terminal j, where,

$$F_{ij}(x_{ij}) = \begin{cases} \max\{M_{ij},\ x_{ij}\} & \text{if } x_{ij} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

The model may be written as:

$$\text{Minimize} \quad \sum_{(ij)\in\mathcal{S}} C_{ij}[F_{ij}(x_{ij})y_{ij} + v_{ij}] + \sum_{i\in\mathcal{T}}[C_i^B x_i^B + C_i^E(y)w_i] \tag{27}$$

$$\text{subject to} \quad \sum_{j\in\mathcal{T}} r_{ij}^d = 1 \text{ and } \{r_{ij}^d\} \in \mathcal{L}, \quad i,j,d \in \mathcal{T} \tag{28}$$

$$x_{ij}^d = \left[w_{id} + \sum_{k\in\mathcal{T}} x_{ki}^d \right] r_{ij}^d, \quad i,j,d \in \mathcal{T} \tag{29}$$

$$r_{ij}^d \leq y_{ij}, \quad i,j,d \in \mathcal{T} \tag{30}$$

$$\sum_{j\in\mathcal{T}} v_{ij} - \sum_{k\in\mathcal{T}} v_{ki} = w_i, \quad i \in \mathcal{T} \tag{31}$$

$$w_i = \sum_{k\in\mathcal{T}} F_{ki}(x_{ki}) - \sum_{j\in\mathcal{T}} F_{ij}(x_{ij}), \quad i \in \mathcal{T} \tag{32}$$

$$y_{ij} \in \{0,1\}, \quad (i,j) \in \mathcal{S} \tag{33}$$

$$r_{ij}^d \in \{0,1\}, \quad (i,j) \in \mathcal{S} \tag{34}$$

$$x_{ij}^d,\ v_{ij}^d,\ \hat{x}_{ij} \geq 0, \quad i,j,d \in \mathcal{T} \tag{35}$$

The objective function (27) computes the total cost of dispatching trailers according to the determined service level, moving the loaded and empty trailers, and handling freight in terminals. Constraints (28) and (29) ensure that freight itineraries obey routing restrictions and that demand is satisfied. Relation (30) is the usual linking constraint that ensures that only operated services are used. Equations (31) and (32) balance the empty flows.

The modeling framework is strongly influenced by the LTL context and the considerable challenges associated with the large size of the LTL carriers operating at the national level in the United States. It may be viewed as an extension of the arc-based multicommodity network design formulation ((4) - (9)) in Section 4.1, with no explicit capacities and a number of complicating constraints. The authors implemented a heuristic procedure based on the hierarchical decomposition of the problem into a master problem and several subproblems. The master problem is a simple network design problem where the total system cost (27) is computed for each given configuration of selected services. The design is modified by adding or dropping one arc at a time (Powell, 1986a). Each time the design is modified, the subproblems must be solved and the objective function must be evaluated. The first subproblem concerns the routing of loaded LTL trailers and it

is solved by shortest-path-type procedures with tree constraints (Powell and Koskosidis 1992). The empty balancing subproblem is solved as a minimum cost transshipment formulation with adjusted supply and demand to account for timing conditions not included in the original formulation (Roy and Delorme 1989, use a similar approach).

The model and solution method are at the core of an interactive decision support system, dubbed APOLLO (Advanced Planner Of LtL Operations), and has been implemented at a major U.S. LTL carrier. Impressive results are reported with respect to the impact of the system both on load planning operations and strategic studies of potential terminal location. Powell and Scheffi (1986, 1989) present in more details the functionalities of APOLLO and discuss its performances on actual problems. They also emphasize the importance of allowing planners to interact with the software to explore alternatives and to select among various options. In this way, planners are better positioned to understand how the system works and, ultimately, to accept its suggestions. The same modeling framework was also used as the basis for the development of a more comprehensive load planning system called SYSNET (Braklow *et al.* 1992), implemented at one of the largest LTL carriers in the United States. In this version, the issue of running direct services, bypassing breakbulk terminals, was explicitly addressed by including such services into the service network. The routing of the freight also acknowledged the geographic and labor structure of the company and considered the relay points where trailers are passed from one driver to the next. The resulting network representation is huge. Heuristics based on company operating rules are used to prune it before the optimization routines are called upon. Other than the optimization model and procedures, the planning system includes demand forecasting, database management, user monitoring and control functionalities. The system has been used with great success to build the load plan, to study the location and dimension of breakbulks, to determine the routing of loaded and empty trailers, and to study which directs should be added or dropped.

5.3 Deterministic Dynamic Service Network Design

When schedules are contemplated, a *time* dimension must be introduced into the formulation. This is usually achieved by representing the operations of the system over a certain number of *time periods* by using a *space-time* network.

The representation of the physical network is replicated in each period. Starting from its origin in a given period, a service arrives (and leaves, in the case of intermediary stops) later at other terminals. Services thus generate temporal service links, between different terminals at different time periods. Temporal links that connect two representations of the same terminal at two different time periods may represent the time required by terminal activities or the freight waiting for the next departure. The costs associated with the arcs of this network are similar to those used in the static formulations of the previous subsections. Additional arcs may be used to capture penalties for arriving too

early or too late.

There are again two types of decision variables. Integer design variables are associated with each service. Restricted to $\{0, 1\}$ values, these variables indicate whether or not the service leaves at the specified time. When several departures may take place in the same time period, general (nonnegative) integer variables must be used. (Note that one can always use $\{0, 1\}$ variables only by making the time periods appropriately small.) Continuous variables are used to represent the distribution of the freight flows through this service network.

The resulting formulations are network design models similar to those presented in Section 4, but on a significantly larger network due to the time dimension. Actually, any of the two previous modeling frameworks, service network design with frequency variables or derived output, may be used as the basis for a dynamic scheduling model. The sheer size of the dynamic network, as well as the additional constraints usually required by the time dimension, makes this class of problems even harder to solve than the static ones. Thus, the pioneering effort of Morlok and Peterson (1970), which integrated blocking, train formation, and train scheduling into a very large mixed integer formulation, never did yield a solution method or an application. Heuristic methods have been used so far.

Farvolden and Powell (1991, 1994) present a dynamic service network design model for LTL transportation. The formulation allows for several departures in the same period, but the simpler $\{0, 1\}$ version is solved. An efficient primal-partitioning with column generation algorithm is used to solve the freight routing problem for any given service configuration (Farvolden, Powell, and Lustig 1992). This was also used to determine the dual variables for service links used to develop an add-drop heuristic for the design problem. The methodology appeared interesting, especially concerning the quality of the evaluation of the add and drop moves. No comprehensive experimental analysis is available, however. Equi *et al.* (1997) determine which shipments of a given good are to be performed and the schedules of the vehicles that will undertake them. The model is a mixed-integer formulation. The proposed heuristic decomposes the problem according to a Lagrangian-type decomposition and proceeds in two steps: a metaheuristic implements tabu search ideas to approximate the design subproblem, and a transportation problem addresses the scheduling part. The methodology has been successfully applied to the problem of transporting wood from cutting areas to ports.

Haghani (1989) attempts to combine the empty car distribution with the train make-up and routing problems. The dynamic network includes normal and express modes for each service route for each time period, but traffic on each link is pre-specified and access to express links is restricted to given markets. Travel times are fixed. Linear functions are used for costs and delays, except for classification, which makes use of a convex congestion function. The dynamic service network design has continuous empty and loaded car flows and integer engine flows. A heuristic decomposition approach is used

to solve somewhat simpler problems and appears efficient for small rail systems. The study also points to better performances, in terms of operating costs, of an integrated formulation as compared to the "traditional" hierarchical approach.

Gorman (1998a) also attempts to integrate the various service network design aspects into a scheduled operating plan that minimizes operating costs, meets the customer's service requirements, and obeys the operation rules of a particular railroad. Model simplifications must be introduced in order to achieve a comprehensive mathematical network design formulation. The solution method is innovative. A hybrid metaheuristic, a tabu-enhanced genetic search, is used to generate candidate train schedules, which are evaluated on their economic, service, and operational performances. On relatively small but realistic problems, the metaheuristic performed very well. A major U.S. railroad has successfully used this model for strategic scenario analysis of their operations (Gorman, 1998b). This work emphasizes the interesting perspectives offered by modern heuristics in addressing complex service network design problems.

5.4 More Service Network Design Models

Several other service network design modeling efforts make use of $\{0, 1\}$ mixed integer network flow formulations similar to the network design models in Section 4. Keaton (1989, 1991, and 1992) proposes a model to develop operating plans for railroads. The model aims to determine which pairs of terminals to connect by direct service, and whether to offer more than one train a day, as well as the routing of freight and the blocking of rail cars. The service network is made up of one network for each origin-destination pair of terminals in the system with positive demand. Links represent trains and connections in yards, as well as *a priori* determined blocking alternatives. Continuous car flows and integer train connections represent the decision variables. All cost functions are linear – there is no congestion and fixed average yard delays are used. The model minimizes the total cost computed as the sum of fixed train costs, car time-related costs, and classification costs. The maximum number of blocks that may be built in a yard yields the linking constraints. Feasibility constraints limit the maximum number of connections and the minimum number of trains for a given pair of terminals. Solutions were obtained by using a Lagrangian relaxation of the linking constraints combined with various heuristics based on operation rules. Results were mixed. While the model was used to perform a number of analyses on relatively small systems, convergence difficulties were also reported.

Newton (1996), Newton, Barnhart, and Vance (1998), and Barnhart, Jin, and Vance (2000) also address the rail blocking problem. They formulate it as a network design problem, where nodes correspond to classification yards, and candidate blocks correspond to arcs. No fixed costs are associated with blocks, but several capacity restrictions are introduced to limit the number of blocks and the total volume of freight processed at

each yard. The first two references present a path-formulation and a branch-and-price solution approach (Barnhart *et al.* 1998). In the third paper, a dual-based Lagrangian relaxation is used to decompose the problem into easier-to-address subproblems: a continuous multicommodity flow problem and an integer block formulation that selects blocks that satisfy the yard capacity constraints. Subgradient optimization is used to solve the Lagrangian dual, while column generation is applied to the flow subproblem. To solve the block subproblem, a branch-and-cut approach is used, where constraints that force the connectivity of at least one path for each commodity are added to the nodes of the enumeration tree. With these constraints, the Lagrangian relaxation identifies a better bound than previously. By incorporating significant data preprocessing to reduce the number of potential blocks and paths, the method could address the problem of a major American railroad and propose blocking plans that represent significant cost improvements.

Kuby and Gray (1993) developed an early model for the design of the network of an express package delivery firm. It is a path-based $\{0, 1\}$ network design model, similar to formulation (13) – (17), where multistop aircraft routes must be selected in and out of a given hub. Paths were generated *a priori*, and the model was solved with a standard mixed-integer package. Analyses illustrated the cost effectiveness of a design with multiple stops over a pure hub-and-spoke network. Kim, Barnhart, and Ware (1999) propose more comprehensive models for the design of the multimodal version of the problem (Barnhart and Schneur, 1996, address a simplified version of the problem). Here, several hubs and aircraft types are considered, while trucks may perform pickup and delivery activities, as well as transportation over limited distances. The problem is further complicated by *time window* restrictions on pickup and delivery times at major collection centers, as well as on the sorting periods at hubs. One product is considered in the application. The authors examine arc, path, and tree-based formulations, and select the latter since it significantly reduces the size of the problem. To solve the linear relaxation of the resulting formulation, the authors combine heuristics to further reduce the size of the problem, cut-set inequalities, and column generation. Branch-and-bound is then used to obtain an integer solution. The paper by Kim and Barnhart (1997) presents a good summary of the authors' experience with these difficult problems and the branch-and-price-and-cut methodology.

The design of postal networks and services forms a class of problems very close to those just mentioned. The LTL frequency service network design by Roy (1984) has already been applied to the design of express letter services for Canada Post. The reorganization of the German postal services belongs to the same problem class, albeit on a more comprehensive scale. To bring the problem down to manageable proportions, Grünert and Sebastian (2000; see also Grünert, Sebastian, and Thärigen 1999 and Buedenbender, Grünert, and Sebastian 2000) decompose it into several subproblems: the optimization of the night airmail network, the design of the groundfeeding and delivery transportation system, the scheduling of operations. Vehicle routing models and techniques are used

for the routing and scheduling tasks. A discrete dynamic network design formulation, similar to those discussed in Section 5.3, is also proposed. The air network design formulation is further decomposed into a direct flight problem and a hub system problem; both yield fixed cost, multicommodity, capacitated network design formulations with side constraints. To optimize these formulations, the authors propose combinations of classical heuristics, tabu search and evolutionary metaheurstics, and exact mathematical programming methods (e.g., branch-and-bound). A decision support system integrates the models and associated solution methods, as well as the tools required to handle the data, models, and methods, and to assist the decision process.

Armacost, Barnhart, and Ware (2002) also address next-day express(air) delivery service design but through a different methodological approach. The authors transform the problem formulation by defining variables that represent combinations of service routes. The new variables implicitly account for the flow distribution and thus yield a pure design formulation for which stronger bounds and thus more efficient solution methods may be derived. The results obtained on data from a major U.S. express shipment firm is very encouraging and emphasizes the need to continue to explore the network design formulations for new insights and more efficient solution methods.

6 Operational Planning and Management

The ultimate goal of any transportation firm is to make profits and improve, or at least maintain, its competitive position. To this end, strategic and tactical plans can be drawn up to guide operations, but the operational capabilities of the firm will ultimately determine its performance.

There are many different issues that must be addressed at the operational level in order to ensure that demand is satisfied within the required service criteria and that the resources of the carrier are efficiently used. Most of these issues must consider the *time* factor. For example, an empty truck must be assigned and moved following a customer request; empty rail cars have to be repositioned, otherwise, soon, idle equipment will be observed at some terminals while others will not be able to satisfy demands; a container must arrive in time to be loaded on the departing ship; a truck has to pick up a load within a specified time window; and so on and so forth. For other types of operations, the very notion of a planned solution does not make sense and the whole operation must continuously adapt and react in real time. Consider, for example, truckload motor carrier services where drivers learn their next assignment only after the current task is concluded. Thus, the need to answer customer requests in real time, to conform to time restrictions on operations, and to integrate in today's decisions their possible impact on future decisions and performances, emphasize the *dynamic* aspect of operational planning and management issues for a freight carrier.

Many models traditionally used in transportation planning use known static data as their input. Tactical planning formulations, for example, consider aggregated forecast demand data as "known". However, the real world in which these models are used is in a constant state of change and solutions cannot always be implemented as planned. If traffic is slower than predicted, vehicles may arrive late at customers' locations or at the terminal. Forecasted customer requests for empty containers or trailers may not materialize while unexpected demands may have to be satisfied. The planned supplies of empty vehicles at depots may thus be unsettled and additional empty movements may have to be performed. Consequently, the dynamic aspect of operations is further compounded by the *stochasticity* inherent to the system, that is, by the set of uncertainties that are characteristic of real-life management and operations. Increasingly, these characteristics are reflected in the models and methods aimed at operational planning and management issues, as illustrated in this section.

6.1 Crew Scheduling

Crews are assigned to vehicles and convoys in order to support the planned operations. There are also numerous other issues related to manpower management such as the scheduling of reserve crews, terminal employees (e.g., Nobert and Roy 1998), maintenance crews, etc. A significant body of methodological and technological knowledge has been developed to deal with these issues, especially in the context of transit (bus and passenger rail) and airline transportation. Some form of *set covering* model is generally used. The resulting mixed-integer formulation is usually very large and it is addressed by column generation and branch-and-price techniques. See, for example, Barnhart and Talluri (1997), Desrosiers *et al.* (1995), Desaulniers *et al.* (1998ab).

These methodologies were developed for applications where detailed schedules are known and adhered to. Consequently, although a few similar developments have targeted crew scheduling issues in the freight transportation industry (e.g., Crainic and Roy 1990, 1992), currently it appears that better results can be achieved by applying the class of methodologies used to dynamically allocate resources to tasks described in Section 6.4.

6.2 Terminal and Linehaul Operations

Terminal and line managers, operators, and dispatchers face a host of control and dispatch issues that form the subject of an extensive literature. The corresponding models and methods aim either to analyze and plan operations or to assist the real-time dispatch of resources and control of operations. A brief enumeration of a number of important issues and references follows.

Terminal models mainly address issues related to the estimation of delays associated with the various operations: load or unload freight, classify vehicles, form blocks and trains, transfer freight between vehicles or convoys, etc. The restricted number of available resources and the large volumes of freight and vehicles that require service result in congestion conditions usually evaluated through the average (and, sometimes, the variance) of the associated waiting time. Delays may also result from the need to wait for the planned connection. Queuing formulations are generally used to derive models for these phenomena (e.g., Crainic 1988). Petersen (1977bc; see also Petersen 1971ab and Petersen and Fullerton 1975) presented what is probably the first comprehensive analysis of yard delays. Bulk service queues, where service is performed for groups of customers (cars) emerge as the main methodological approach. They are difficult to solve, however, in all but the simplest limiting cases. Turnquist and Daskin (1982) use similar formulations for their rail yard model but relax a number of restrictions in order to obtain a more tractable model. Daskin and Walton (1983) propose a set of queuing models to represent the lightering operations (transfers from large ocean tankers to smaller vessels) in crude transportation.

Powell (1981, 1986b; see also Powell and Humblet, 1986) undertook a significant study of bulk queues and their applications to modeling delays in transportation terminals. He proposed efficient numerical methods (Powell, 1986d) and closed-form approximations (Powell, 1986c) to compute the moments of the distributions. Closed-form approximation formulas have also been proposed by Crainic and Gendreau (1986). Such closed-form approximations of delays in freight terminals (as well as on the lines of the systems) are equally important as generators of functions and measures for the service network design and the strategic planning models presented in Sections 5 and 3, respectively. A different perspective on yard-blocking performance is offered by Daganzo (1987ab). Based on direct analyses of the departure schedules, policies, and operational rules, formulae are determined for a number of performance measures - number of tracks in the yard and number of switches per car, for example - for various blocking strategies.

PORTS? Raymond??

Many rail line models aim to represent the delays that result when trains meet (on single-track lines) or when one train overtakes another. When traffic volumes are low, analytical formulae may be obtained directly from the corresponding operating rules (Petersen, 1974, 1975ab). Queuing models are again the methodology of choice, the approach being similar to that used to analyze rail yard operations when congestion conditions occur (Petersen, 1977c). More recently, Chen and Harker (1990) and Harker and Hong (1990) consider the case when services are scheduled and evaluate the mean and variance of delays on double and single-track lines, respectively. Hallowell and Harker (1996) evaluate and predict performance on partial double-track rail line with scheduled traffic.

The preceding models may be combined and, eventually, approximated, to yield formulations that may be used in more comprehensive planning systems. Petersen and Taylor (1982), Petersen (1984), Crainic, Ferland, and Rousseau (1984), and Crainic, Florian, and Léal (1990) integrate queuing submodels or functions into the planning systems they propose.

A different line of research relative to rail line models addresses issues related to the scheduling and pacing of trains on a line. Jovanović and Harker (1991) propose a mixed integer formulation, solved by branch-and-bound, to assist the tactical (weekly or monthly) scheduling of trains on a line. The model is embedded in the SCAN1 software. The issue of optimally pacing trains over a line is addressed by Kraay, Harker, and Chen (1991), Higgins, Ferreira, and Kozan (1995) and Higgins, Kozan, and Ferreira (1996, 1997). Network-based mixed integer formulations also appear here. The application of genetic, tabu search, and hybrid metaheuristics to the same problem is explored by Higgins, Ferreira, and Kozan (1997).

6.3 Empty Vehicle Distribution and Repositioning

A particularly important and challenging issue for freight carriers is the need to move empty vehicles. Indeed, the geographic differences in demand and supply for each commodity type often result in an accumulation of empty vehicles in regions where they are not needed and in deficits of vehicles in other regions that require them. Then, vehicles must be moved empty, or additional loads must be found, in order to bring them where they will be needed to satisfy known and forecasted demand in the following planning periods. This operation is known as *repositioning* and is a major component of what is known as *fleet management*. In its most general form, fleet management covers the whole range of planning and management issues from procurement of power units and vehicles to vehicle dispatch and scheduling of crews and maintenance operations. Often, however, the term designates a somewhat restricted set of activities: allocation of vehicles to customer requests and repositioning of empty vehicles.

Moving vehicles empty does not directly contribute to the profit of the firm but it is essential to its continuing operations. Consequently, one attempts to minimize empty movements within the limits imposed by the demand and service requirements. *Empty balancing*, the distribution of empty vehicles to balance the supply and demand in future periods, is a major objective of dispatchers and a central component of planning and operations of most transportation firms. This issue must also be considered at the tactical level. In rail transportation, for example, empty rail cars are put on the same trains as loaded ones and thus contribute to an increase in the number of trains, in the volume of vehicles handled in terminals and, ultimately, in system costs and delays. For planning purposes, the demand for empty cars may be approximated and introduced in tactical model by viewing empties as another commodity to be transported (e.g., Crainic,

Ferland, and Rousseau 1984). A similar approach may also be used for the planning of multimodal regional or national systems (e.g., Crainic, Florian, and Léal 1990). The issue is also relevant in LTL trucking where empty balancing is an integral part of a transportation plan. In this case, a load plan is first obtained for the actual traffic demands, and an empty balancing model is then solved to reposition the empties (see Delorme and Roy 1989 and Braklow *et al.* 1992, for example).

Numerous studies reflect the significant research and development effort that has been dedicated to empty vehicle distribution issues. Interested readers may start exploring this field with the review of Dejax and Crainic (1987). It includes contributions going back to the '60s and spans the whole spectrum of modeling approaches from simple static transport models to formulations that integrate the dynamic and stochastic characteristics of the problem. In the following, we recall some of the main articles and models in this field.

The first empty vehicle allocation models used straightforward transportation formulations (e.g., Leddon and Wrathall 1967; Misra 1972; Baker 1977). Given estimations of future supply and demand of empty cars of a homogeneous fleet at the yards of the network, and the cost in car-hours usually, for each pair of yards, the distribution of empty cars is optimized to minimize the total cost.

A significant step forward in modeling capabilities was achieved with the explicit consideration of the time perspective. A space-time diagram represents the various paths that vehicles may travel to reach their proper destination at a specified time (Figure 3 illustrates such a network). The resulting formulation takes the form of a deterministic dynamic transshipment network model, where flows are optimized such that either the total cost is minimized, or the profitability of the system is maximized. Starting with the pioneering contributions of White (1968) and White and Bomberault (1969) for rail car distribution, and of White (1972) for container allocation, many models that aimed for the distribution of empty vehicles, took the form of a dynamic transshipment network optimization problem (e.g., Herren 1973, 1977; McGaughey, Gohring, and McBrayer 1973). Linear programming and network flow algorithms were usually applied. This line of research is still very active today. The formulations are more complex, though. Multiple commodities, substitutions, integer flows, are some of the characteristics that add realism to these formulations (Shan 1985; Chih 1986; Turnquist and Markowicz 1989; Markowicz and Turnquist 1990; Turnquist 1994; and others). Alternatively, the strict schedules and booking policies enforced by many European railways impose additional conditions on empty vehicle distribution, such as limited hauling capacity for empties, and pre-defined itineraries (Joborn 1995; Holmberg, Joborn, and Lundgren 1998; Joborn *et al.* 2001).

The explicit consideration of uncertainties in empty vehicle distribution models constitutes another significant methodological contribution. The first comprehensive effort

in this direction was made by Jordan and Turnquist (1983) for rail. The formulation aims to maximize the profits of the firm, and integrates revenues from performing the service as well as various costs from moving cars between yards, holding them at yards, or from not filling orders due to stockouts. The model structure is again a multicommodity, dynamic network. Stochasticity of supply, demand, and travel times is explicitly considered. The resulting model is a nonlinear optimization formulation, solved by using the Frank-Wolfe algorithm (1956). A similar approach is proposed by Beaujon and Turnquist (1991) for a model that simultaneously considers vehicle inventories at terminals and their allocation in order to answer fleet-sizing issues. The whole research area addressing the dynamic allocation of limited resources in uncertain environments naturally continues these important developments.

6.4 Dynamic Allocation of Resources

Many operational problems, fleet management in particular, dynamically allocate limited resources to requests and tasks. For example, empty vehicles, trailers and rail cars are allocated to the appropriate terminals; motive power tractors and locomotives to services; crews to vehicles or services; loads to driver-truck combinations; empty containers from depots to customers and returning containers from customers to depots; and so on. All these problems have several common characteristics:

1. Some future demands are known, but most can only be forecasted, and unpredictable requests may happen.
2. Many requests materialize in real or quasi-real time and must be acted upon in relatively short time.
3. Once a resource is allocated to an activity, it is no longer available for a certain duration (whose length may be subject to variations as well).
4. Once a resource becomes available again, it is often in a different location than its initial one.
5. The value of an additional unit of a given resource at a location greatly depends on the total quantity of resources available (which are determined from previous decisions at potentially all terminals in previous periods) and the current demand.

This is an extremely rich field both for research and development and for applications. In a sense, it extends and complements the empty vehicle distribution problems described previously. The latest developments in the field also allow to plan and control the activities of several resources simultaneously (Powell 1996b, 1998; Powell and Carvalho 1998b; Powell and Shapiro, 2001). Dynamic and stochastic network formulations

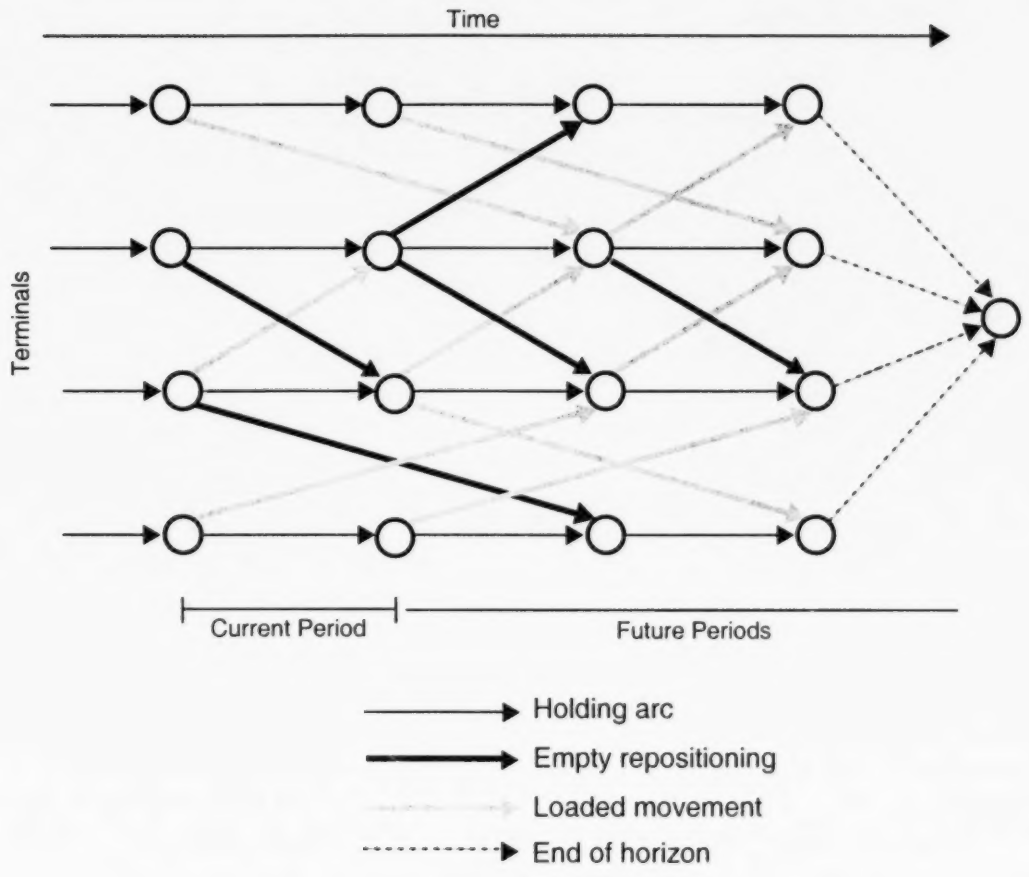

Figure 3: Dynamic Network Representation of Operations

have been, and continue to be, extensively studied for these problems. This has resulted in important modeling and algorithmic results. A number of these results have been transferred to industry (Powell *et al.* 1992, for example). The interested reader should consult the excellent synthesis and review by Powell, Jaillet, and Odoni (1995) and the numerous references quoted in this work. In the following, we briefly illustrate two main modeling approaches.

One may represent dynamic allocation issues by an activity graph similar to the one displayed in Figure 3. Here, the operations of a simple four-terminal system are schematically drawn for a certain length of time, which is arbitrarily divided into three periods. At each terminal, there are a number of vehicles that are available to satisfy customer requests during the current period and in future ones. Customer demands have precise characteristics, such as the origin and destination of movement, and pickup

and delivery dates (with time window restrictions, eventually). At any period, a vehicle may be assigned to a customer demand in the current period and at the current location, moved to another location to satisfy a known future request, held at the current location, or moved empty to another location in preparation for future, forecasted demands.

Accepting requests and performing the corresponding movements implies expenses and generates revenues. Crainic, Gendreau, and Dejax (1993) developed a model for the assignment and management of a heterogeneous fleet of containers where loaded movements are exogenously accepted. Here, the objective is to minimize the total operating cost, including substitutions and stockouts. Several other models also address the issue of whether a request is profitable with respect to the operation of the system and should therefore be accepted. Indeed, repositioning empty vehicles does not generate any immediate revenues. One may be ready to incur these expenses, however, in the hope that, as a consequence, vehicles will be adequately posted to take advantage of future (known, forecasted or estimated) opportunities. Refused requests represent lost business opportunities, while accepted but unsatisfied ones generally result in penalties. Powell *et al.* (1992) and Powell (1996a) present such applications to truckload motor carrier transportation.

A classical modeling approach for this class of problems is to consider the entire planning horizon with the objective of maximizing the *total system profit* computed as the sum of the profit resulting from decisions taken for the current period, plus the expected profit over future periods. The usual constraints apply: satisfy the demand; do not use more than the number of available vehicles; adhere to specific operations rules; etc. When the state of the system and its environment in future periods is known, or assumed to be known, the resulting formulation is deterministic and is often written as a network flow optimization model with additional constraints.

The major difficulty with this approach becomes apparent when the uncertainties in future demands, as well as, eventually, uncertainties related to performing the operations, are explicitly considered. In this case, decisions taken "now" for future periods cannot be based on sure data, but only on estimations of how the system will evolve, which demand will materialize, and so on. From a mathematical programming point of view, random variables are used to represent the stochastic elements and decisions in future periods. Consequently, the expectation of future profits that appears in the objective function of the model becomes a very complex, recursive stochastic equation where the statistical expectation of the total profit must be computed over all possible realizations of all random variables.

To address this complex issue, the model generally takes the form of a recourse formulation. Such formulations are based on the idea that today's decisions are taken within today's deterministic context but using an estimation of the variability of the random factors, and that their consequences are reflected in later decisions. The recourse represents

these later decisions which must be taken to adjust the initial policies once the actual realization of the random variables is observed. In the simplest possible recourse formulation, called *simple recourse*, it is assumed that one does not attempt to change the decisions but pays a penalty when the observed value of a random variable is different from the estimation. More complex formulations, such as nodal, tree, and network recourse attempt to evaluate the possible modifications to the initial decisions, and the impact on the total expected profit. Refer to Powell (1988), Frantziskakis (1990), Powell and Frantzeskakis (1994), and Powell and Cheung (1994ab) for details. Powell (1987), Frantziskakis and Powell (1990), Cheung and Powell (1996a), Chen and Powell (1999), Powell and Cheung (2000) extend the recourse methodology and present increasingly more complex and precise methods to approximate the recourse function of multistage, dynamic, stochastic networks. An excellent analysis of the application of these approaches to the dynamic fleet management problems for truckload motor carriers, as well as a discussion of the merits and difficulties of stochastic formulations, may be found in Powell (1996a). Cheung and Powell (1996b) further compare these approaches in the context of dynamic distribution problems. Cheung and Chen (1998) apply the same type of methodology to the problem of distributing empty containers in an international maritime system.

These formulations, which are generally difficult to solve, also make use of various criteria to discretize, aggregate, and end time. For example, in Figure 3, the theoretically infinite future planning and operation horizon has been reduced to three periods. When the recourse formulation is solved, the periods could be further aggregated, all future periods being considered as one; this corresponds to a two-period formulation, as opposed to n-period, otherwise. Then, in actual applications, the models are used in a rolling horizon environment where, as time advances, a new period is added at the end of the horizon. An important issue is then how to approximate what happens in all the periods beyond the artificially fixed end of the horizon, and how to integrate this approximation into the recourse function. Powell, Jaillet, and Odoni (1995) present an excellent review of this class of formulations.

A different approach recently championed by Powell (1995), Powell *et al.* (1995), and Powell and Carvalho (1998a; see also Carvalho 1996; Carvalho and Powell 2000) addresses resource allocation problems as *Logistic Queueing Networks, LQN*. In this case, at each node of the time-space diagram there are two queues: one of resources and one of tasks requesting resources. Figure 4 illustrates a possible configuration for two terminals over two periods. Two "resources" are managed, vehicles and loads, and their levels currently known or approximated at each terminal are displayed. Available vehicles may be allocated to loads already at terminals. Arrows illustrate other possible actions: move loaded vehicles from one terminal to another, where they will increase the inventory of empty vehicles; hold empty vehicles for use in subsequent periods; move empty vehicles to reposition them at a different terminal; determine where to send a vehicle that becomes empty and what vehicles from which terminal to assign to new loads. The objective is to maximize the total profit generated by operating the system to satisfy demand.

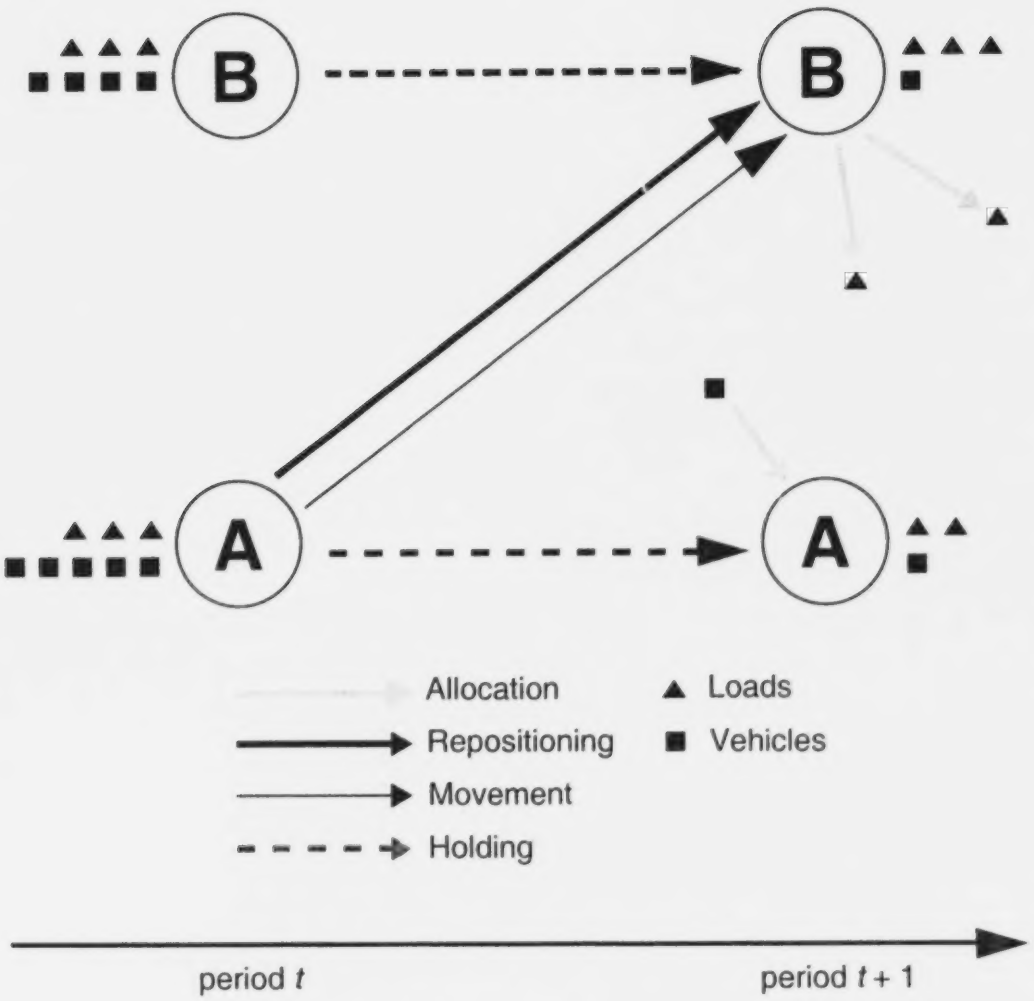

Figure 4: Logistics Queuing Network

The basic idea of the *LQN* methodology is to cast the formulation as a recursive dynamic model and to decompose the resulting optimization problem for each period into "easy-to-solve" local subproblems. In the applications described, each subproblem corresponds to the assignment of vehicles to tasks (loaded or empty movements, for example) at a given terminal. But, in order to evaluate the worth of allocating a vehicle to a loaded or repositioning movement, one has to know, or evaluate, not only the operating costs and the profit of the load, but also the value of the empty vehicle at the destination terminal. Furthermore, the dynamics of the system make this value depend on future decisions at all terminals. At each period and for each vehicle type, these values are approximated by measuring how desirable it is to have one more vehicle at each terminal. The resulting *potentials* are then used to build a linear approximation of the part of the recursive objective function that corresponds to future periods. Gradients of this approximated objective functions with respect to the supply of vehicles at terminals are used to adjust the potentials, as well as the upper limits on empty movements.

The general solution approach proceeds iteratively in a series of forward and backward passes along the time axis. At each iteration, the forward pass assigns vehicles to tasks, the backward pass computes gradients, and a control adjustment phase modifies the potentials and the bounds on empty movements. The process continues until "convergence" is ensured. The latest developments (Godfrey and Powell 2002a,b) use nonlinear approximations and present truly impressive results for realistic fleet management applications.

The LQN approach appears to offer a very interesting framework for a wide variety of real situations that may be efficiently represented and solved. It offers, in particular, a rather straightforward way to explicitly take into account various considerations, such as time windows, labor restrictions, and substitutions, by addressing them at the level of the local subproblem. The application of LQN methodology to the real-time management of fleets of containers and flatcars for intermodal operations presented by Powell and Charvalho (1998b) offers very encouraging results both in terms of actual results (significant savings in operation costs are forecasted) and of further applications to resource allocation problems and service network design models.

7 Perspectives

We have presented a number of major issues, models, and methodologies in long distance freight transportation planning and management. Many significant methodological advances have been achieved and many have been successfully transferred to actual practice. However, many research opportunities and challenges still exist.

The advent of *Intelligent Transportation Systems (ITS)* will have a tremendous im-

pact on the planning and operations of freight transportation. ITS technologies increase the flow of available data, improve the timeliness and quality of information, and offer the possibility to control and coordinate operations in real-time. Significant research efforts are required to adequately model the various planning and management problems under ITS and real-time information, and to develop efficient solution methods. Some efforts have already been undertaken relative to the real-time dispatching, assignment, routing, and re-routing of vehicles (Regan, 1997; Regan, Mahmassani, and Jaillet, 1995, 1996ab, 1998; Yang, Mahmassani, and Jaillet, 1998; Yang, Jaillet, and Mahmassani (2002); Gendreau *et al.*, 1996, 1998, 1999; Gendreau and Potvin, 1998) and the study of the impact of new technologies on the planning and performance of intermodal classification yards (Bostel and Dejax, 1998).

The rapid and sustained development of the *electronic business* way of interacting with customers and partners is already modifying how transportation firms plan and operate. In many respects, e-business and ITS are related and the the challenges associated to the real-time response mentioned above are also encountered here. E-business also brings (or should bring) easier access to loads through various e-marketplaces. Many of these offer increasingly sophisticated auction mechanisms to determine the allocation of loads and the associated prices. The fleet management models and tools have to integrate these possibilities. A major challenge is related to determining on what loads to bid and the bidding strategy, in particular when loads that would combine in interesting routes must be negotiated separately e.g., (Abrache, Crainic, and Gendreau 2001, Chang, Crainic, and Gendreau 2002, Crainic and gendreau 2002). It is difficult at this time to adequately predict the whole extend of impact of ITS and e-business on transportation science theory, methods, and practice, but we are convinced that it will be major and comprehensive.

The study of network design formulations and solution methods still offers considerable challenges; from a theoretical point of view, of course, but also when contemplating applications to huge problem instances with very large number of commodities. The same may be said of dynamic and stochastic formulations. In fact, one observes that more and more formulations explicitly consider the dynamic and stochastic characteristics of the problems under study. The trend may be observed not only for issues traditionally associated with actual operations, but also for problems considered "tactical", such as load planning and service network design. Generally speaking, however, the literature does not offer trusted solution methods capable of addressing schedule (dynamic) service network design problems of realistic dimensions and complexity. The study of the formulations and their properties (e.g., reformulations, bounds, cuts) should be continued. A number of decomposition ideas (according to the time period or node, for example) have also been advanced and are worth investigating. Such approaches will also present "natural" parallelization characteristics that should facilitate the implementation of efficient solution methods.

Metaheuristics play an increasingly important role in obtaining good solutions to dif-

ficult problems within reasonable computing times. Work is still needed, however, to develop more efficient and more robust procedures and to better understand the conditions under which each approach performs best. Hybrids, combining characteristics of two or more metaheuristics, offer interesting, but challenging perspectives.

Parallel and distributed computation offers another challenging perspective with potentially great rewards: to solve realistically modeled and dimensioned problem instances within reasonable times. Each class of problems and algorithms presents its own challenges. The parallel exploration of branch-and-bound trees, and the collaborative search undertaken by several metaheuristics or by metaheuristics and exact methods, are only two exciting research areas. Parallel computing also offers the possibility of designing an architecture to efficiently answer complex requests in real, or quasi-real time. These ideas that have just begun to be considered (e.g. Séguin *et al.*, 1997), have a great potential for the development of intelligent and efficient decision support tools for ITS and other real-time transportation systems.

Acknowledgments

I desire to take this opportunity to express my gratitude toward my colleagues and friends, Professors Gilbert Laporte and Bernard Gendron, who read earlier versions of this paper. Their comments have helped improve the presentation significantly. I also desire to thank Ms. Lucie Larin for the professional drawing of the figures.

Financial support for this project was provided by N.S.E.R.C. (Canada) and the Fonds F.C.A.R. (Québec).

Annex: Location models

Discrete location models may generally be cast as particular cases of network design formulations (see, for example, Magnanti and Wong 1984; Nemhauser and Wolsey 1988; Ahuja, Magnanti, and Orlin 1993). Yet, relatively few people work on both types of problems and the extensive literature for each type remains fairly separate. The historic development of the two domains, rooted in different applications, is largely the source of this phenomenon. To simplify the presentation, we follow tradition and present the two classes of formulations separately, and attempt to bring about a unified view from an integrated notation and cross-references.

7.1 Location models

Location problems involve the siting of one or several facilities, usually at vertices of a network, in order to facilitate the movement of goods or the provision of services along the network. Demand ("customers") is also usually present at the vertices of the network. These may be the same or different from the vertices where facilities may be located. The main location models are often classified as follows (see Crainic and Laporte 1997):

1. *Covering models.* Locate facilities at the vertices of a network so that the demand vertices are covered by a facility, i.e. they lie within a given distance of a facility. The coverage distance, usually related to the shortest path distance between the facility and the demand nodes, may be the same for all vertices, or may depend on the specific facility and demand points. The problem can be to minimize the cost of locating facilities, subject to a constraint stating that all vertices are covered. If one operates within a fixed budget, then an objective can be to maximize the demand covered by the facilities. See Shilling, Jayaraman, and Barkhi (1993) for a recent survey of these models.

2. *Center models.* Locate p facilities at vertices of a network in order to minimize the maximum distance between a demand point and a facility. See Handler (1990) for a review of these formulations.

3. *Median models.* Locate p facilities at vertices on the network and allocate demand to these facilities in order to minimize the total weighted distance between facilities and demand points. If facilities are uncapacitated and p is fixed, one obtains the so-called p-median problem. In such a case, each vertex is assigned to its closest facility. If p is a decision variable and facilities are uncapacitated, this defines the *Uncapacitated Plant Location Problem* (UPLP); if facilities are capacitated, one obtains the *Capacitated Plant Location Problem* (CPLP). Labbé, Peeters, and Thisse (1995) and Daskin (1995) review these models.

Covering problems are typically associated with the location of public facilities such as health clinics, post offices, libraries, and schools. Center problems often arise when establishing the location of emergency facilities such as fire or ambulance stations. Median problems are directly relevant to logistics service design and freight distribution. Here we will describe a general CPLP formulation as well as two extensions particularly suited to the type of issues we address in this chapter. For a review of these and other models, see Mirchandani and Francis (1990), Daskin (1995), Drezner (1995), or Labbé, Peeters, and Thisse (1995). Labbé and Louveaux (1997) present an annotated bibliography concerning discrete location problems.

We define the *multicommodity* formulation of the CPLP in terms of a network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ represents the set of *nodes* or *vertices* where known demand exists and facilities (plants, warehouses, depots, terminals, intermodal platforms) may be implemented to satisfy them. The set of *arcs* or *links* $\mathcal{A} = \{a = (i,j) |\ i,j \in \mathcal{N},\ i \neq j\}$ includes all the possible ways to move directly (no intermediate nodes) between two nodes in $\mathcal{N}$. The set $\mathcal{P}$ includes the *products* or *commodities* that may move on the network. Let i and j be node indices and p the product index, and define:

f_j: cost of locating a facility at vertex j;

d_i^p: demand for commodity p at node i;

c_{ij}^p: transportation cost per unit of flow of commodity p between locations i and j;

u_j: capacity of a facility located at j;

y_j: logical variable equal to 1 if and only if a facility is located at j, and 0 otherwise;

x_{ij}^p: quantity of the demand of product p at node i served by a facility located at j.

The model is then:

$$\text{Minimize} \quad \sum_{j \in \mathcal{N}} f_j y_j + \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} c_{ij}^p x_{ij}^p \tag{36}$$

$$\text{subject to} \quad x_{ij}^p \leq min\{u_j, d_i^p\} y_j \qquad i,j \in \mathcal{N},\ p \in \mathcal{P} \tag{37}$$

$$\sum_{j \in \mathcal{N}} x_{ij}^p = d_i^p \qquad i \in \mathcal{N},\ p \in \mathcal{P} \tag{38}$$

$$\sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{N}} x_{ij}^p \leq u_j y_j \qquad j \in \mathcal{N} \tag{39}$$

$$y_j = 0 \text{ or } 1 \qquad j \in \mathcal{N} \tag{40}$$

$$x_{ij}^p \geq 0 \qquad i,j \in \mathcal{N},\ p \in \mathcal{P}. \tag{41}$$

In this model, the objective function represents the sum of fixed facility and transportation costs. It is assumed these costs are scaled over the same planning horizon.

Constraint (37) expresses the condition that demand at i can only be served from j if a facility is located at j to the extent of the capacity of the facility, while constraint (38) states that the entire demand of each location must be allocated to facilities. The utilization of $min\{u_j, d_i^p\}$ instead of simply u_j makes for a tighter formulation when $d_i^p < u_j$. Constraint (39) ensures that the capacity of a facility is never exceeded by its assigned demand. In most CPLP formulations, the p index is dropped and the demand, flow variables, and transportation costs are scaled by the corresponding d_i^p factors. Constraint (39) is relaxed in UPLP formulations, which are significantly simpler to solve than capacitated ones.

The previous model is a mixed-integer formulation that may be approached by any of the methodologies available for this class of problems (e.g. Nemhauser and Wolsey, 1988 or Salkin and Mathur, 1989). A widely used methodology is to relax one or several groups of constraints in a Lagrangian fashion to obtain a simpler problem (Geoffrion, 1974). Relaxing capacity constraint (39) constitutes a very attractive alternative since it yields an UPLP-type of problem for which several efficient solution methods exist (DUALOC-based procedures, initially devised by Erlenkotter in 1978, still appear the most effective; Cornuéjols, Nemhauser, and Wolsey 1990 survey solution alternatives for the UPLP). Cornuéjols, Sridharan, and Thizy (1991) study this and many other relaxations for the CPLP (forty-one relaxations that yield only seven different bounds) and compare them in terms of theoretical bound quality, computational precision, as well as the quality of the heuristic feasible solutions they generate.

A sequence of multiplier adjustments and resolutions of the relaxation subproblem yields a lower bound on the optimal value of the original formulation. As for multipliers, they may be adjusted by using a nondifferentiable optimization technique, subgradient or bundle, for example (Lemaréchal 1989). *Dual ascent* is another often-used approach to obtain this lower bound. In this case, the dual formulation of the linear relaxation of the problem is the starting point. Dual variables are then iteratively and systematically increased, while conforming to the complementary slackness conditions. This process yields a lower bound on the optimal value of the original formulation. An upper bound on this optimal value is then obtained as the objective value of a feasible solution heuristically derived from the solution to the relaxed problem. The lower and upper bounds are then usually integrated into an implicit enumeration scheme such as the *branch-and-bound* algorithm.

As for other mixed-integer programming formulations, the polyhedral structure of the model may be studied to derive *valid inequalities* (or *cuts*) to be added to the formulation. A detailed description of this methodology is given in Nemhauser and Wolsey (1988). Briefly, the objective is to construct, or approximate, the convex hull of the mixed-integer programming formulation by adding valid inequalities. If one succeeds and the convex hull is found, the original problem may be solved by linear programming methods. The *cutting plane* method is based on this idea and proceeds via a succession of resolutions

of the linear relaxation of the problem and cut generations. If the convex hull can only be approximated, the bounds may be strengthened, yielding more efficient *branch-and-bound* algorithms. Major contributions to this field with applications to the CPLP come from Leung and Magnanti (1989).

In many cases, the additional complexity introduced to account for the particularities of the application at hand and the large size of the problem instance make the exact resolution of the problem impractical. Heuristics are then used to obtain solutions of, hopefully, good quality. A number of heuristics, e.g. the *greedy* and the *interchange* procedures originally introduced by Kuehn and Hamburger (1963) for the UPLP, aim to avoid mathematical programming techniques altogether but are not very successful for capacitated models. The relaxations and dual-ascent methods presented above are also often used as heuristics with interesting results (see, for example, Klincewicz and Luss, 1986 or Van Roy 1986). Modern heuristics, principally *Tabu Search* (Glover and Laguna 1997), *Simulated Annealing* (Laarhoven and Aarts 1987), and *Genetic Algorithms* (Goldberg 1989), are also increasingly being applied.

The basic formulation may be modified to reflect particular problem classes and applications. *Location-routing* formulations explicitly consider the interplay between the location of terminals or depots, and the pickup and delivery routes that must be built to address the distribution problem. Routing-related models are well beyond the scope of this chapter. See Laporte (1988), Federgruen and Simchi-Levi (1995), and Campbell *et al.* (1998) for more complete descriptions of the subject. Two classes of formulations generalize the CPLP and are of particular interest for planning freight transportation systems: the *production-distribution* models, and the *hub location* models. Both formulations aim to capture the potential economies of scale associated with the *consolidation* of freight (or passengers). In the following, we briefly introduce these formulations. See Daskin (1995) for a more in-depth presentation of these and other extensions and applications.

To formulate the production-distribution model, assume $\mathcal{N}$ is made up of three, not necessarily disjoint, sets of nodes: $\mathcal{F}$ = the sites of production facilities or *plants*; $\mathcal{W}$ = the locations of *warehouses*; and $\mathcal{C}$ = the demand points of *customers*. Commodities are produced at plants, from where they are shipped to customers either directly or via a warehouse.Any combination of plant, warehouse, and customer could be sited at any location of set $\mathcal{N}$. The main issues addressed by the formulation concern the number and location of warehouses, and the flow pattern of products through the system: either directly from plant to customer, or through an intermediate warehouse. Figure 5 illustrates these concepts: seven customers are served from three plants through two warehouses. Plain arrows indicate the corresponding movements. Dashed arrows stand for the direct shipments that are permitted and could be selected. Note that, implicitly, the formulation also addresses issues related to the production planning at each plant, for example, how much to produce and for which customer. We define:

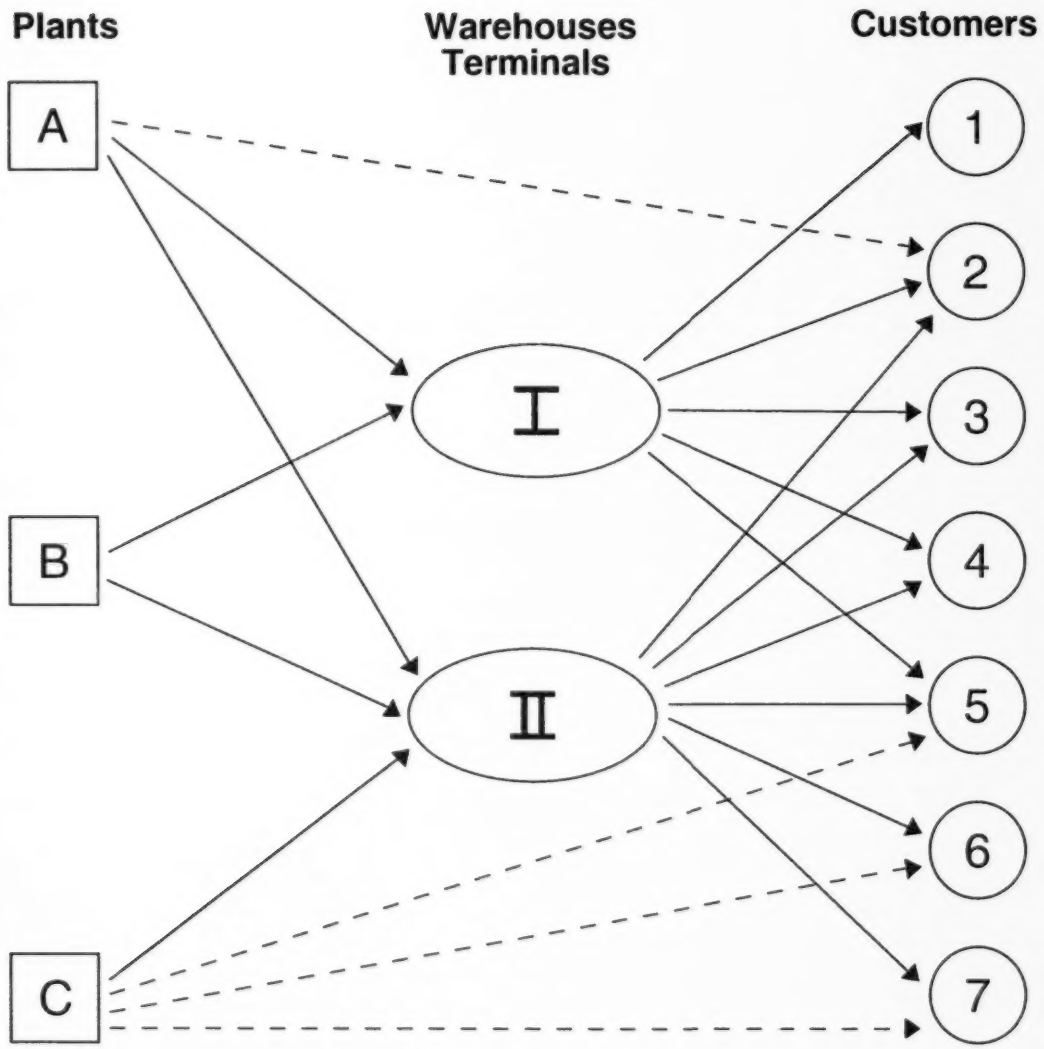

Figure 5: Production-Distribution Network

f_j: cost of locating a warehouse at location $j \in \mathcal{W} \subseteq \mathcal{N}$;

d_i^p: demand for commodity p at customer $i \in \mathcal{C} \subseteq \mathcal{N}$;

c_{kji}^p: transportation cost per unit of flow of commodity $p \in \mathcal{P}$ from plant $k \in \mathcal{F} \subseteq \mathcal{N}$ to customer i transiting through warehouse j;

u_j: capacity of a warehouse located at j;

u_k^p: production capacity of plant k for commodity p – one usually assumes independent commodity production capacities;

y_j: logical variable equal to 1 if and only if a warehouse is located at j, and 0 otherwise;

x_{kji}^p: flow of product p from plant k shipped to customer i through a warehouse at candidate location j.

The model is then:

$$\text{Minimize} \quad \sum_{j \in \mathcal{W}} f_j y_j + \sum_{p \in \mathcal{P}} \sum_{i \in \mathcal{C}} \sum_{j \in \mathcal{W}} \sum_{k \in \mathcal{F}} c_{kji}^p x_{kji}^p \tag{42}$$

$$\text{subject to} \quad x_{kji}^p \leq min\{u_k^p, u_j, u_k^p\} y_j \qquad k \in \mathcal{F},\ j \in \mathcal{W}\ , i \in \mathcal{C},\ p \in \mathcal{P} \tag{43}$$

$$\sum_{p \in \mathcal{P}} \sum_{k \in \mathcal{F}} \sum_{i \in \mathcal{C}} x_{kji}^p \leq u_j y_j \qquad j \in \mathcal{W} \tag{44}$$

$$\sum_{j \in \mathcal{W}} \sum_{i \in \mathcal{C}} x_{kji}^p \leq u_k^p \qquad k \in \mathcal{F},\ p \in \mathcal{P} \tag{45}$$

$$\sum_{k \in \mathcal{F}} \sum_{j \in \mathcal{W}} x_{kji}^p = d_i^p \qquad i \in \mathcal{C},\ p \in \mathcal{P} \tag{46}$$

$$y_j = 0 \text{ or } 1 \qquad j \in \mathcal{W} \tag{47}$$

$$x_{kji}^p \geq 0 \qquad k \in \mathcal{F}\ , i \in \mathcal{C},\ j \in \mathcal{W},\ p \in \mathcal{P}. \tag{48}$$

In this formulation, equation (42) computes the total cost of opening warehouses and distributing the products. Constraints (43) and (44) ensure that only open warehouses are used to distribute products. Relation (44) also enforces the warehouse capacity constraint. Production capacities at plants are enforced by equation (46). The formulation may be enhanced by including partial warehouse capacities for all or some of the commodities considered, and by relaxing the independence assumption on the production and, eventually, warehousing capacities. Many practical considerations may add variables and constraints to the formulation. Of these, issues related to the level of inventory at plants and warehouses are of prime importance. See Aikens (1985) for a review of production-distribution models.

Klincewicz (1990) presents an interesting application of the production-distribution class of formulations to freight transportation problems. Consolidation terminals must

be located (the "warehouses" from the previous formulation) and no capacity is imposed on either the production sites or the terminals. Piecewise linear concave functions are used, however, to capture the economies of scale in transportation cost associated with concentrated flows. When transportation costs are linear between plants and terminals, or terminals and customers, the problem decomposes into concave cost uncapacitated location problems and optimal algorithms are proposed. For the general case, heuristics that solve a series of these optimal formulations are studied. Recently, Pirkul and Jayaraman (1996, 1998) combine Lagrangian relaxations with rules based on the logistics characteristics of the system (costs and distances, principally) to find very good solutions (very low optimality gaps) to a production-distribution problem where capacitated plants and warehouses must be sited to satisfy the customer demands for several products.

Hub-networks represent the organization of many transportation systems: passenger air transport, express mail, less-than-truckload motor carriers, railways, international intermodal container movements, to name just a few prominent ones. In such systems, people or goods are moved from a given origin to a specified destination. To take advantage of economies of scale, the lower demands are moved first to an intermediate point – a *hub* – such as an airport, port, rail yard, or intermodal or consolidation terminal. At a hub, traffic is consolidated and more massive flows are routed to other hubs. Lower frequency services, often operating smaller vehicles, are used in the transportation legs between hubs and terminals. Figure 1 illustrates such a system used in consolidation transportation. Nodes A, B, and C are hubs linked by high frequency and capacity services. Nodes 1 to 9 stand for the origin and destination terminals linked to hubs by feeder services. The figure also emphasizes the possibility for a terminal to be linked to more than one hub, as well as the pickup and delivery operations usually associated to the terminal operations.

Such an organization allows a much higher frequency and quality of service among hubs, and a more efficient utilization of resources. This explains why the airline industry in the U.S. adopted this scheme following its deregulation. The drawback of this type of organization is the increased delays (longer routes and time spent in terminals) experienced by passengers or goods. This explains partly why there is hardly any "pure" hub-and-spoke systems in operation, direct transportation being organized for high demand or high priority origin-destination pairs. The links between terminals 4 and 5, and from hub A to terminal 9 in Figure 1 illustrate this option. Note that smaller firms may take advantage of hub-and-spoke systems and identify profitable niches by offering direct services to markets that large firms serve through hubs.

The location of hubs greatly determines the efficiency and appeal of hub-based systems. Often, these issues are addressed by simulating the impact of various locations by using tactical or operational models of the system. Specific hub-location formulations and algorithms have also been proposed, however. A basic linear mixed-integer formulation (Campbell, 1994a,b) would resemble the previous production-distribution model

with decision variable x_{ij}^{kl}, where:

x_{ij}^{kl} : quantity of demand with origin i and destination j that passes through hubs k and l

This variable explicitly indicates the routing of the flow and the utilization of the hubs; of course, these itineraries could include routes with any number of intermediate hubs, including no hubs at all. See O'Kelly (1987) for a quadratic programming formulation of this problem and Daskin (1995) for a more detailed discussion of the subject. Heuristics have mainly been proposed for this class of formulations 9Klincewicz, 1991 and Skorin-Kapov and Skorin-Kapov, 1994).

The multicommodity location-allocation with balancing requirements model (Crainic *et al.*, 1989) combines a number of characteristics of the model classes described in this section. The formulation aims to determine the best logistics structure of the land distribution and transportation component of an international container shipping company. The main questions concern the selection of inland depots, the assignment of customers to depots for each container type and direction of operations (allocation of empty containers from depots to customers and return of empties from customers to depots), and the determination of the main repositioning flows of empty containers to counter the regional differences between supply and demand. The objective is to minimize the total logistic cost made up of the "fixed" cost of using the terminals (since the company does not build terminals, but rather uses space in rail yards and ports), the container transportation cost between customers and depots, and the long haul container transportation cost among depots. Transportation costs are specific to each container type and mode of transportation. The problem presents characteristics similar to a multicommodity UPLP where discrete choice variables stand for the selection of depots, but with a complete multicommodity network flow structure that is similar to the hub-location formulation that has the allocation and repositioning of containers as continuous decision variables. Dual ascent methods, of DUALOC type, have been proposed (Crainic and Delorme, 1995), as well as branch-and-bound (Crainic, Delorme, and Dejax 1993; Gendron and Crainic, 1995, 1997) and tabu search procedures (Crainic *et al.*, 1993, 1995a,b).

References

Abrache, J., Crainic, T.G., and Gendreau, M. (2001). Design Issues for Multi-object Combinatorial Auctions. In *Proceedings of The Fourth International Conference on Electronic Commerce Research (ICECR-4)*, volume 2, pages 412–423. Cox School of Business, Southern Methodist University, Dallas TX.

Ahuja, R.K. (1997). Flows and Paths. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 283–309. John Wiley & Sons, New York, NY.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993). *Network Flows - Theory, Algorithms, and Applications*. Prentice-Hall, Englewood Cliffs, NJ.

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., and Reddy, M.R. (1995). Applications of Networks Optimization. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, pages 1–83. North-Holland, Amsterdam.

Aikens, C.H. (1985). Facility Location Models for Distribution Planning. *European Journal of Operational Research*, 22:263–279.

Armacost, A.P., Barnhart, C., and Ware, K.A. (2002). Composite Variable Formulations for Express Shipment Service Network Design. *Transportation Science*, 36(1).

Assad, A.A. (1980). Models for Rail Transportation. *Transportation Research A: Policy and Practice*, 14A:205–220.

Baker, L. (1977). Overview of Computer Based Models Applicable to Freight Car Utilization. Report prepared for U.S. Department of Transportation, NTIS, Sprinfield VA, U.S.A.

Balakrishnan, A., Magnanti, T.L., and Mirchandani, P. (1997). Network Design. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 311–334. John Wiley & Sons, New York, NY.

Balakrishnan, A., Magnanti, T.L., and Wong, R.T. (1989). A Dual–Ascent Procedure for Large–Scale Uncapacitated Network Design. *Operations Research*, 37(5):716–740.

Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors (1995). *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*. North-Holland, Amsterdam.

Barnhart, C., Hane, C.A., and Vance, P.H. (2000a). Using Branch-and-Price-and-Cut to Solve Origin-Destination Integer Multicommodity Flow Problems. *Operations Research*, 48(2):318–326.

Barnhart, C., Jin, H., and Vance, P.H. (2000b). Railroad Blocking: A Network Design Application. *Operations Research*, 48(4):603–614.

Barnhart, C., Johnson, E.L., Nemhauser, G.L., Savelsbergh, M.W.F., and Vance, P.H. (1998). Branch-and-Price: Column Generation for Solving Huge Integer Programs. *Operations Research*, 46(3):316–329.

Barnhart, C. and Schneur, R.R. (1996). Network Design for Express Freight Service. *Operations Research*, 44(6):852–863.

Barnhart, C. and Talluri, K. (1997). Airlines Operations Research. In A.E. McGarity and C. ReVelle, editors, *Civil and Environmental Engineering Systems: An Advanced Applications Text*. John Wiley, New York.

Beaujon, G.J. and Turnquist, M.A. (1991). A Model for Fleet Sizing and Vehicle Allocation. *Transportation Science*, 25(1):19–45.

Berger, D., Gendron, B., Potvin, J.-Y., Raghavan, S., and Soriano, P. (2000). Tabu Search for a Network Loading Problem with Multiple Facilities. *Journal of Heuristics*, 6(2):253–267.

Bostel, N. and Dejax, P. (1998). Models and Algorithms for Container Allocation Problems on Trains in a Rapid Transshipment Shunting Yard. *Transportation Science*, 32(4):370–379.

Braklow, J.W., Graham, W.W., Hassler, S.M., Peck, K.E., and Powell, W.B. (1992). Interactive Optimization Improves Service and Performance for Yellow Freight System. *Interfaces*, 22(1):147–172.

Buedenbender, K., Grünert, T., and Sebastian, H.-J. (2000). A Hybrid Tabu Search/Branch and Bound Algorithm for the Direct Flight Network Design Problem. *Transportation Science*, 34(4):364–380.

Camerini, P.M., Fratta, L., and Maffioli, F. (1978). On Improving Relaxation Methods by Modified Gradient Techniques. *Mathematical Programming Study*, 3:26–34.

Campbell, A., Clarke, L., Kleywegt, A., and Savelsberg, M. (1998). The Inventory Routing Problem. In T.G. Crainic and G. Laporte, editors, *Fleet Management and Logistics*, pages 95–113. Kluwer Academic Publishers, Norwell, MA.

Campbell, J.F. (1994a). Integer Programming Formulations of Discrete Hub Location Problem. *European Journal of Operational Research*, 72:387–405.

Campbell, J.F. (1994b). A Survey of Network Hub Location Problems. *Studies in Locational Analysis*, 6:31–49.

Carvalho, T.A. (1996). *A New Approach to Solving Dynamic Resource Allocation Problems.* PhD thesis, Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ, U.S.A.

Carvalho, T.A. and Powell, W.B. (2000). A Multiplier Adjustment Method for Dynamic Resource Allocation Problems. *Transportation Science*, 34(2):150–164.

Cascetta, E. (2001). *Transportation Systems Engineering: Theory and Methods.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Chang, T.S., Crainic, T.G., and Gendreau, M. (2001). Designing Advisors for Fleet Management and E-Auctions. Technical report, Centre de recherche sur les transports, Université de Montréal, Canada.

Chen, B. and Harker, P.T. (1990). Two Moments Estimation of the Delay on a Single-Track Rail Line with Scheduled Traffic. *Transportation Science*, 24(4):261–275.

Chen, Z.-L. and Powell, W.B. (1999). A Convergent Cutting-planne and Partial-sampling Algorithm for Multistage Linear Programs with Recourse. *Journal of Optimization Theory and Applications*, 103(3):497–524.

Cheung, R.K. and Chen, C.-Y. (1998). A Two-Stage Stochastic Network Model and Solution Methods for the Dynamic Empty Container Allocation Problem. *Transportation Science*, 32(2):142–162.

Cheung, R.K.-M. and Powell, W.B. (1996a). An Algorithm for Multistage Dynamic Networks with Random Arc Capacities, with an Application to Dynamic Fleet Management. *Operations Research*, 44(6):951–963.

Cheung, R.K.-M. and Powell, W.B. (1996b). Models and Algorithms for Distribution Problems with Uncertain Demands. *Transportation Science*, 30(1):43–59.

Chih, K. C.-K. (1986). *A Real Time Dynamic Optimal Freight Car Management Simulation Model of the Multiple Railroad, Multicommodity, Temporal Spatial Network Flow Problem.* PhD thesis, Princeton University, Princeton, NJ, U.S.A.

Chouman, M., Crainic, T.G., and Gendron, B. (2001). Revue des inégalités valides pertinentes aux problèmes de conception de réseaux. Publication CRT-2001-38, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Chouman, M., Crainic, T.G., and Gendron, B. (2002). Valid Inequalities for Multicommodity Caparitated Fixed Charge Network Design. Technical report, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Cordeau, J.-F., Toth, P., and Vigo, D. (1998). A Survey of Optimization Models for Train Routing and Scheduling. *Transportation Science*, 32(4):380–404.

Cornuéjols, G., Nemhauser, G.L., and Wolsey, L.A. (1990). The Uncapacitated Facility Location Problem. In Michandani, P.B. and Francis, R.L., editors, *Discrete Location Theory*, pages 119–171. John Wiley & Sons, New York, NY.

Cornuéjols, G., Sridharan, R., and Thizy, J.M. (1991). A Comparison of Heuristics and Relaxations for the Capacitated Plant Location Problem. *European Journal of Operational Research*, 50:280–297.

Crainic, T.G. (1982). *Un modèle de plannification tactique pour le transport ferroviaire des marchandises.* PhD thesis, Université de Montréal, Montréal, QC, Canada.

Crainic, T.G. (1984). A Comparison of Two Methods for Tactical Planning in Rail Freight Transportation. In J.P. Brans, editor, *Operational Research'84*, pages 707–720. North-Holland, Amsterdam.

Crainic, T.G. (1988). Rail Tactical Planning: Issues, Models and Tools. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 463–509. Springer-Verlag, Berlin.

Crainic, T.G. (2002). Parallel Computation, Co-operation, Tabu Search. In Rego, C. and Alidaee, B., editors, *Adaptive Memory and Evolution: Tabu Search and Scatter Search.* Kluwer Academic Publishers, Norwell, MA.

Crainic, T.G., Dejax, P.J., and Delorme, L. (1989). Models for Multimode Multicommodity Location Problems with Interdepot Balancing Requirements. *Annals of Operations Research*, 18:279–302.

Crainic, T.G. and Delorme, L. (1995). Dual-Ascent Procedures for Multicommodity Location-Allocation Problems with Balancing Requirements. *Transportation Science*, 27(2):90–101.

Crainic, T.G., Delorme, L., and Dejax, P.J. (1993a). A Branch-and-Bound Method for Multicommodity Location with Balancing Requirements. *European Journal of Operational Research*, 65(3):368–382.

Crainic, T.G., Dufour, G., Florian, M., and Larin, D. (1999). Path Analysis in STAN. In C. Zopounidis and D. Despotis, editors, *Proceedings 5th International Conference of the Decision Sciences Institute*, pages 2060–2064. New Technologies Publications, Athens, Greece.

Crainic, T.G., Dufour, G., Florian, M., and Larin, D. (2002). Path Recovery/Reconstruction and Applications in Nonlinear Multimodal Multicommodity Networks. In Gendreau, M. and Marcotte, P., editors, *Transportation and Network Analysis - Current Trends.* Kluwer Academic Publishers, Norwell MA.

Crainic, T.G., Ferland, J.-A., and Rousseau, J.-M. (1984). A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, 18(2):165–184.

Crainic, T.G., Florian, M., Guélat, J., and Spiess, H. (1990a). Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System. *Transportation Research Record*, 1283:97–124.

Crainic, T.G., Florian, M., and Larin, D. (1994). STAN: New Developments. In Al S. Khade and R. Brown, editors, *Proceedings of the 23rd Annual Meeting of the Western Decision Sciences Institute*, pages 493–498. School of Business Administration, California State University, Stanislaus CA.

Crainic, T.G., Florian, M., and Léal, J.-E. (1990b). A Model for the Strategic Planning of National Freight Transportation by Rail. *Transportation Science*, 24(1):1–24.

Crainic, T.G., Frangioni, A., and Gendron, B. (2001). Bundle-Based Relaxation Methods for Multicommodity Capacitated Network Design. *Discrete Applied Mathematics*, 112:73–99.

Crainic, T.G. and Gendreau, M. (1986). Approximate Formulas for the Computation of Connection Delays under Capacity Restrictions in Rail Freight Transportation. In *Research for Tomorrow's Transport Requirements*, pages 1142–1155. Fourth World Conference on Transport Research, Vancouver, Canada. Vol. 2.

Crainic, T.G. and Gendreau, M. (2002a). Cooperative Parallel Tabu Search for Capacitated Network Design. *Journal of Heuristics*.

Crainic, T.G. and Gendreau, M. (2002b). Freight Exchanges and Carrier Operations: Issues, Models, and Tools. Technical report, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Crainic, T.G., Gendreau, M., and Dejax, P.J. (1993b). Dynamic Stochastic Models for the Allocation of Empty Containers. *Operations Research*, 43:102–302.

Crainic, T.G., Gendreau, M., and Farvolden, J.M. (2000). A Simplex-Based Tabu Search Method for Capacitated Network Design. *INFORMS Journal on Computing*, 12(3):223–236.

Crainic, T.G., Gendreau, M., Soriano, P., and Toulouse, M. (1993c). A Tabu Search Procedure for Multicommodity Location/Allocation with Balancing Requirements. *Annals of Operations Research*, 41:359–383.

Crainic, T.G. and Laporte, G. (1997). Planning Models for Freight Transportation. *European Journal of Operational Research*, 97(3):409–438.

Crainic, T.G. and Nicolle, M.-C. (1986). Planification tactique du transport ferroviaire des marchandises: quelques aspects de modélisation. In *Actes du Premier Congrès International en France de Génie Industriel*, pages 161–174. CEFI-AFCET-GGI, Paris. Tome 1.

Crainic, T.G. and Rousseau, J.-M. (1986). Multicommodity, Multimode Freight Transportation: A General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research B: Methodology*, 20B:225–242.

Crainic, T.G. and Roy, J. (1988). O.R. Tools for Tactical Freight Transportation Planning. *European Journal of Operational Research*, 33(3):290–297.

Crainic, T.G. and Roy, J. (1990). Une approche de recouvrement d'ensembles pour l'établissement d'horaires des chauffeurs dans le transport routier de charges partielles. *R.A.I.R.O. - Recherche opérationnelle*, 24(2):123–158.

Crainic, T.G. and Roy, J. (1992). Design of Regular Intercity Driver Routes for the LTL Motor Carrier Industry. *Transportation Science*, 26(4):280–295.

Crainic, T.G. and Toulouse, M. (2002). Parallel Strategies for Meta-heuristics. In F. Glover and G. Kochenberger, editors, *State-of-the-Art Handbook in Metaheuristics*. Kluwer Academic Publishers, Norwell, MA.

Crainic, T.G., Toulouse, M., and Gendreau, M. (1995a). Parallel Asynchronous Tabu Search for Multicommodity Location-Allocation with Balancing Requirements. *Annals of Operations Research*, 63:277–299.

Crainic, T.G., Toulouse, M., and Gendreau, M. (1995b). Synchronous Tabu Search Parallelization Strategies for Multicommodity Location-Allocation with Balancing Requirements. *OR Spektrum*, 17(2/3):113–123.

Crowder, H. (1976). Computational Improvements for Subgradient Optimization. In *Symposia Mathematica*, volume XIX. Academic Press, London.

Daganzo, C.F. (1987a). Dynamic Blocking for Railyards: Part I. Homogeneous Traffic. *Transportation Research B: Methodological*, 21B(1):1–27.

Daganzo, C.F. (1987b). Dynamic Blocking for Railyards: Part II. Heterogeneous Traffic. *Transportation Research B: Methodological*, 21B(1):29–40.

Daskin, M.S. (1995). *Network and Discrete Location. Models, Algorithms, and Applications*. John Wiley & Sons, New York, NY.

Daskin, M.S. and Walton, C.M. (1983). An Approximate Analytic Model of Supertanker Lightering Operations. *Transportation Research B: Methodological*, 17B(3):201–219.

Dejax, P.J. and Crainic, T.G. (1987). A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science*, 21(4):227–247.

Delorme, L., Roy, J., and Rousseau, J.-M. (1988). Motor-Carrier Operation Planning Models: A State of the Art. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 510–545. Springer-Verlag, Berlin.

Desaulniers, G., Desrosiers, J., Gamache, M., and Soumis, F. (1998a). Crew Scheduling in Air Transportation. In T.G. Crainic and G. Laporte, editors, *Fleet Management and Logistics*, pages 169–185. Kluwer Academic Publishers, Norwell, MA.

Desaulniers, G., Desrosiers, J., Ioachim, I., Solomon, M.M., Soumis, F., and Villeneuve, D. (1998b). A Unified Framework for Deterministic Time Constrained Vehicle Routing and Crew Scheduling Problems. In T.G. Crainic and G. Laporte, editors, *Fleet Management and Logistics*, pages 57–93. Kluwer Academic Publishers, Norwell, MA.

Desrosiers, J., Dumas, Y., Solomon, M.M., and Soumis, F. (1995). Time Constrained Routing and Scheduling. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 35–139. North-Holland, Amsterdam.

Drezner, Z., editor (1995). *Facility Location. A Survey of Applications and Methods.* Springer-Verlag, New York, NY.

Dror, M., editor (2000). *Arc Routing: Theory, Solutions and Applications.* Kluwer Academic Publishers, Norwell, MA.

Epstein, R. (1998). *Linear Programming and Capacitated Network Loading.* PhD thesis, Sloan School of Management, Massachusetts Institute of Technology.

Equi, L., Gallo, G., Marziale, S., and Weintraub, A. (1997). A Combined Transportation and Scheduling Problem. *European Journal of Operational Research*, 97(1):94–104.

Erlenkotter, D. (1978). A Dual–Based Procedure for Uncapacitated Facility Location. *Operations Research*, 26:992–1009.

Farvolden, J.M. and Powell, W.B. (1991). A Dynamic Network Model for Less-Than-Truckload Motor Carrier Operations. Working Paper 90-05, Department of Industrial Engineering, University of Toronto, Toronto, ON, Canada.

Farvolden, J.M. and Powell, W.B. (1994). Subgradient Methods for the Service Network Design Problem. *Transportation Science*, 28(3):256–272.

Farvolden, J.M., Powell, W.B., and Lustig, I.J. (1992). A Primal Partitioning Solution for the Arc–Chain Formulation of a Multicommodity Network Flow Problem. *Operations Research*, 41(4):669–694.

Federgruen, A. and Simchi-Levi, D. (1995). Analysis of Vehicle Routing and Inventory-Routing Problems. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 297–373. North-Holland, Amsterdam.

Florian, M. and Hearn, D. (1995). Networks Equilibrium Models and Algorithms. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network*

Routing, volume 8 of *Handbooks in Operations Research and Management Science*, pages 485–550. North-Holland, Amsterdam.

Franck, M. and Wolfe, P. (1956). An Algorithm for Quadratic Programming. *Naval Research Logistics Quaterly*, 3:95–110.

Frantzeskakis, L. (1990). *Dynamic Networks with Random Arc Capacities, with Application to the Stochastic Dynamic Vehicle Allocation Problem*. PhD thesis, Department of Civil Engineering and Operation Research, Princeton University, Princeton, NJ, U.S.A.

Frantzeskakis, L. and Powell, W.B. (1990). A Successive Linear Approximation Procedure for Stochastic Dynamic Vehicle Allocation Problems. *Transportation Science*, 24(1):40–57.

Friesz, T.L., Gottfried, J.A., and Morlok, E.K. (1986). A Sequential Shipper-Carrier Network Model for Predicting Freight Flows. *Transportation Science*, 20:80–91.

Friesz, T.L., Tobin, R.L., and Harker, P.T. (1983). Predictive Intercity Freight Network Models. *Transportation Research A: Policy and Practice*, 17A:409–417.

Friesz, T.L. and Harker, P.T. (1985). Freight Network Equilibrium: A Review of the State of the Art. In A.F. Daughety, editor, *Analytical Studies in Transport Economics*, chapter 7. Cambridge University Press, Cambridge.

Gendreau, M., Badeau, P., Guertin, F., Potvin, J.-Y., and Taillard, É.D. (1996). A Solution Procedure for Real-Time Routing and Dispatching of Commercial Vehicles. In *Third Annual World Congress on Intelligent Transportation Systems*. distributed on CD.

Gendreau, M., Guertin, F., Potvin, J.-Y., and Séguin, R. (1998). Neighborhood Search Heuristics for a Dynamic Vehicle Dispatching Problem with Pick-Ups and Deliveries. Publication CRT-98-10, Centre de recherche sur les transports, Université de Montréal.

Gendreau, M., Guertin, F., Potvin, J.-Y., and Taillard, É.D. (1999). Tabu Search for Real-Time Vehicle Routing and Dispatching. *Transportation Science*, 33(4):381–390.

Gendreau, M. and Potvin, J.-Y. (1998). Dynamic Vehicle Routing and Dispatching. In T.G. Crainic and G. Laporte, editors, *Fleet Management and Logistics*, pages 115–126. Kluwer Academic Publishers, Norwell, MA.

Gendron, B. and Crainic, T.G. (1994). Relaxations for Multicommodity Network Design Problems. Publication CRT-965, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Gendron, B. and Crainic, T.G. (1994a). Parallel Branch-and-Bound Algorithms: Survey and Synthesis. *Operations Research*, 42(6):1042–1066.

Gendron, B. and Crainic, T.G. (1994b). Parallel Implementations of Bounding Procedures for Multicommodity Capacitated Network Design Problems. Publication CRT-94-45, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Gendron, B. and Crainic, T.G. (1995). A Branch-and-Bound Algorithm for Depot Location and Container Fleet Management. *Location Science*, 3(1):39–53.

Gendron, B. and Crainic, T.G. (1996). Bounding Procedures for Multicommodity Capacitated Network Design Problems. Publication CRT-96-06, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Gendron, B. and Crainic, T.G. (1997). A Parallel Branch-and-Bound Algorithm for Multicommodity Location with Balancing Requirements. *Computers & Operations Research*, 24(9):829–847.

Gendron, B., Crainic, T.G., and Frangioni, A. (1998). Multicommodity Capacitated Network Design. In Sansó, B. and Soriano, P., editors, *Telecommunications Network Planning*, pages 1–19. Kluwer Academic Publishers, Norwell, MA.

Geoffrion, A.M. (1974). Lagrangean Relaxation for Integer Programming. *Mathematical Programming Study*, 2:82–114.

Ghamlouche, I., Crainic, T.G., and Gendreau, M. (2002a). Cycle-based Neighbourhoods for Fixed-Charge Capacitated Multicommodity Network Design. *Operations Research*.

Ghamlouche, I., Crainic, T.G., and Gendreau, M. (2002b). Path Relinking, Cycle-based Neighbourhoods and Capacitated Multicommodity Network Design. *Annals of Operations Research*.

Glover, F. and Laguna, M. (1993). Tabu search. In Reeves, C., editor, *Modern Heuristic Techniques for Combinatorial Problems*, pages 70–150. Blackwell Scientific Publications, Oxford.

Glover, F. and Laguna, M. (1997). *Tabu Search*. Kluwer Academic Publishers, Norwell, MA.

Godfrey, G.A. and Powell, W.B. (2002a). An Adaptive Dynamic Programming Algorithm for Dynamic Fleet Management I: Single Period Travel Times. *Transportation Science*, 36(1).

Godfrey, G.A. and Powell, W.B. (2002b). An Adaptive Dynamic Programming Algorithm for Dynamic Fleet Management II: Multiperiod Travel Times. *Transportation Science*, 36(1).

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.

Golden, B.L. and Assad, A.A., editors (1988). *Vehicle Routing: Methods and Studies*. North-Holland, Amsterdam.

Gorman, M.F. (1998a). An Application of Genetic and Tabu Searches to the Freight Railroad Operating Plan Problem. *Annals of Operations Research*, 78:51–69.

Gorman, M.F. (1998b). Santa Fe Railway Uses an Operating-Plan Model to Improve Its Service Design. *Interfaces*, 28(4):1–12.

Grötschel, M., Monma, C.L., and Stoer, M. (1995). Design of Survivable Networks. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Models*, volume 7 of *Handbooks in Operations Research and Management Science*, pages 617–672. North-Holland, Amsterdam.

Grünert, T. and Sebastian, H.-J. (2000). Planning Models for Long-haul Operations of Postal and Express Shipment Companies. *European Journal of Operational Research*, 122:289–309.

Grünert, T., Sebastian, H.-J., and Thäerigen, M. (1999). The Design of a Letter-Mail Transportation Network by Intelligent Techniques. In Sprague, R., editor, *Proceedings Hawaii International Conference on System Sciences 32*.

Guélat, J., Florian, M., and Crainic, T.G. (1990). A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows. *Transportation Science*, 24(1):25–39.

Haghani, A.E. (1989). Formulation and Solution of Combined Train Routing and Makeup, and Empty Car Distribution Model. *Transportation Research B: Methodological*, 23B(6):433–431.

Hallowell, S.F. and Harker, P.T. (1996). Predicting On-Time Line-Haul performance in Scheduled Railroad Operations. *Transportation Science*, 30(4):364–378.

Handler, G.Y. (1990). p-Center Problems. In Michandani, P.B. and Francis, R.L., editors, *Discrete Location Theory*, pages 305–348. John Wiley & Sons, New York, NY.

Harker, P.T. (1987). *Predicting Intercity Freight Flows*. VNU Science Press, Utrech, The Netherlands.

Harker, P.T. (1988). Issues and Models for Planning and Regulating Freight Transportation Systems. In Bianco, L. and Bella, A. L., editors, *Freight Transport Planning and Logistics*, pages 374–408. Springer-Verlag, Berlin.

Harker, P.T. and Friesz, T.L. (1986a). Prediction of Intercity Freight Flows I: Theory. *Transportation Research B: Methodological*, 20B(2):139–153.

Harker, P.T. and Friesz, T.L. (1986b). Prediction of Intercity Freight Flows II: Mathematical Formulations. *Transportation Research B: Methodological*, 20B(2):155–174.

Harker, P.T. and Hong, S. (1990). Two Moments Estimation of the Delay on a Partially Double-Track Rail Line with Scheduled Traffic. *Journal of the Transportation Research Forum*, 31(1):38–49.

Herren, H. (1973). The Distribution of Empty Wagons by Means of Computer: An Analytical Model for the Swiss Federal Railways (SSB). *Rail International*, 4(1):1005–1010.

Herren, H. (1977). Computer Controlled Empty Wagon Distribution on the SSB. *Rail International*, 8(1):25–32.

Higgins, A., Ferreira, L., and Kozan, E. (1995). Modeling Single-Line Train Operations. *Transportation Research Record*, 1489:9–16.

Higgins, A., Ferreira, L., and Kozan, E. (1997a). Heuristic Techniques for Single Line Train Scheduling. *Journal of Heuristics*, 3(1):43–63.

Higgins, A., Kozan, E., and Ferreira, L. (1996). Optimal Scheduling of Trains on a Single Line Track. *Transportation Research B: Methodological*, 30B:147–161.

Higgins, A., Kozan, E., and Ferreira, L. (1997b). Modelling the Number and Location of Sidings on a Single Line Track Railway. *COmputers & Operations Research*, 3:209–220.

Hiriart-Urruty, J.B. and Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, Berlin.

Hoffman, K.L. and Padberg, M. (1993). Solving Airline Crew-Scheduling Problems by Branch-and-Cut. *Management Science*, 39:657–682.

Holmberg, K., Joborn, M., and Lundgren, J.T. (1998). Improved Empty Freight Car Distribution. *Transportation Science*, 32(2):163–173.

Holmberg, K. and Hellstrand, J. (1998). Solving the Uncapacitated Network Design Problem by a Lagrangian Heuristic and Branch-and-Bound. *Operations Research*, 46(2):247–259.

Holmberg, K. and Yuan, D. (1998). A Lagrangean Approach to Network Design Problems. Report LiTH-MAT-R-1998-01, Department of Mathematics, Linköping Institute of Technology.

Holmberg, K. and Yuan, D. (2000). A Lagrangean Heuristic Based Branch-and-Bound Approach for the Capacitated Network Design Problem. *Operations Research*, 48(3):461–481.

Hurley, W.J. and Petersen, E.R. (1994). Nonlinear Tariffs and Freight Network Equilibrium. *Transportation Science*, 28(3):236–245.

Isard, W. (1951). Interregional and Regional Input-Output Analysis: A Model of a Space-Economy. *The review of Economics and Statistics*, 33:318–328.

Joborn, M. (1995). *Empty Freight Car Distribution at Swedish Railways - Analysis and Optimization Modeling.* PhD thesis, Division of Optimization, Department of Mathematics, University of Linköping, Linköping, Sweden.

Joborn, M., Crainic, T.G., Gendreau, M., Holmberg, K., and Lundgren, J.T. (2001). Economies of Scale in Empty Freight Car Distribution in Scheduled Railways. Publication CRT-2001-12, Centre de recherche sur les transports, Université de Montréal, Montréal, QC, Canada.

Jordan, W.C. and Turnquist, M.A. (1983). A Stochastic Dynamic Network Model for Railroad Car Distribution. *Transportation Science*, 17:123–145.

Jovanović, D. and Harker, P.T. (1991). Tactical Scheduling of Rail Operations: the SCAN I Decision Support System. *Transportation Science*, 25(1):46–64.

Keaton, M.H. (1989). Designing Optimal Railroad Operating Plans: Lagrangian Relaxation and Heuristic Approaches. *Transportation Research B: Methodological*, 23B(6):415–431.

Keaton, M.H. (1991). Service–Cost Tradeoffs for Carload Freight Traffic in the U.S. Rail Industry. *Transportation Research A: Policy and Practice*, 25A(6):363–374.

Keaton, M.H. (1992). Designing Optimal Railroad Operating Plans: A Dual Adjustment Method for Implementing Lagrangian Relaxation. *Transportation Science*, 26:263–279.

Kim, D. and Barnhart, C. (1997). Transportation Service Network Design: Models and Algorithms. Report, Center for Transportation Studies, Massachusetts Institute of Technology.

Kim, D., Barnhart, C., Ware, K., and Reinhardt, G. (1999). Multimodal Express Package Delivery: A Service Network Design Application. *Transportation Science*, 33(4):391–407.

Klincewicz, J.G. (1990). Solving a Freight Transport Problem Using Facility Location Techniques. *Operations Research*, 38(1):99–109.

Klincewicz, J.G. (1991). Heuristics for the p-Hub Location Problem. *European Journal of Operationsl Research*, 53:25–37.

Klincewicz, J.G. and Luss, H. (1986). A Lagrangean Relaxation Heuristic for Capacitated Facility Location with Single Source Constraints. *Journal of the Operational Research Society*, 37:495–500.

Kraay, D., Harker, P.T., and Chen, B. (1991). Optimal Pacing of Trains in Freight Railroads: Model Formulation and Solution. *Operations Research*, 39(1):82–99.

Kuby, M.J. and Gray, R.G. (1993). The Hub Network Design Problem with Stopovers and Feeders: The Case of Federal Express. *Transportation Research A: Policy and Practice*, 27A(1):1–12.

Kuehn, A.A. and Hamburger, M.J. (1990). A Heuristic Program for Locating Warehouses. *Management Science*, 9:643–666.

Laarhoven, P. and Aarts, E.H.L. (1987). *Simulated Annealing: Theory and Applications*. Reidel, Dordrecht.

Labbé, M., Peeters, D., and Thisse, J.-F. (1995). Location on Networks. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 551–624. North-Holland, Amsterdam.

Labbé, M. and Louveaux, F.V. (1997). Location Problems. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 261–281. John Wiley & Sons, New York, NY.

Lamar, B.W., Sheffi, Y., and Powell, W.B. (1990). A Capacity Improvement Lower Bound for Fixed Charge Network Design Problems. *Operations Research*, 38(4):704–710.

Laporte, G. (1988). Location-Routing Problems. In Golden, B.L. and Assad, A.A., editors, *Vehicle Routing: Methods and Studies*, pages 163–197. North-Holland, Amsterdam.

Larin, D., Crainic, T.G., Simonka, G., James-Lefebvre, L., Dufour, G., and Florian, M. (2000). *STAN User's Manual, Release 6*. INRO Consultants, Inc., Montréal, QC, Canada.

Leddon, C.D. and Wrathall, E. (1967). Scheduling Empty Freight Car Fleets on the Louisville and Nashville Railroad. In *Second International Symposium on the Use of Cybernetics on the Railways*, pages 1–6.

Lemaréchal, C. (1989). Nondifferentiable Optimization. In Nemhauser, G.L., Rinnoy Kan, A.H.G., and Todd, M.J., editors, *Optmization*, volume 1 of *Handbooks in Operations Research and Management Science*, pages 529–572. North-Holland, Amsterdam.

Leung, J.M.Y. and Magnanti, T.L. (1989). Valid Inequalities and Facets of the Capacitated Facility Location Problems. *Mathematical Programming*, 44:271–291.

Magnanti, T.L., Mirchandani, P., and Vachani, R. (1993). The Convex Hull of Two Core Capacitated Network Design Problems. *Mathematical Programming*, 60:233–250.

Magnanti, T.L., Mirchandani, P., and Vachani, R. (1995). Modeling and Solving the Two-Facility Capacitated Network Loading Problem. *Operations Research*, 43:142–157.

Magnanti, T.L. and Wong, R.T. (1986). Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, 18(1):1–55.

Markowicz, B.P. and Turnquist, M.A. (1990). Applying the LP Solution to the Daily Distribution of Freight Cars. Technical report, Cornell University.

McGaughey, K., Gohring, W., and McBrayer, R.N. (1973). Planning Locomotive and Caboose Distribution. *Rail International*, 3:151–158.

Minoux, M. (1986). Network Synthesis and Optimum Network Design Problems: Models, Solution Methods and Applications. *Networks*, 19:313–360.

Mirchandani, P.S. and Francis, R.L., editors (1990). *Discrete Location Theory*. John Wiley & Sons, New York, NY.

Misra, S. (1972). Linear Programming of Empty Wagon Disposition. *Rail International*, 3:151–158.

Morlok, E.K. and Peterson, R.B. (1970). A Final Report on a Development of a Geographic Transportation Network Generation and Evaluation Model. In *Proceedings of the Eleventh Annual Meeting*, pages 99–103. Transportation Research Forum.

Nagurney, A. (1993). *Network Economics - A Variational Inequality Approach*. Kluwer Academic Publishers, Norwell, MA.

Nemhauser, G.L. and Wolsey, L.A. (1988). *Integer and Combinatorial Optimization*. Wiley, New York, NY.

Newton, H.N. (1996). *Network Design Under Budget Constraints with Application to the Railroad Blocking Problem*. PhD thesis, Industrial and Systems Engineering, Auburn University, Auburn, Alabama, U.S.A.

Newton, H.N., Barnhart, C., and Vance, P.H. (1998). Constructing Railroad Blocking Plans to Minimize Handling Costs. *Transportation Science*, 32(4):330–345.

Nobert, Y. and Roy, J. (1998). Freight Handling Personnel Scheduling at Air Cargo Terminals. *Transportation Science*, 32(3):295–301.

O'Kelly, M.E. (1987). A Quadratic Integer Program for the Location of Interacting Hub Facilities. *European Journal of Operational Research*, 32:393–404.

Pallottino, S. and Scutellà, M.G. (1998). Shortest path Algorithms in Transportation Models: Classical and Innovative Aspects. In Marcotte, P. and Nguyen, S., editors, *Equilibrium and Advanced Transportation Modelling*, pages 245–281. Kluwer Academic Publishers, Norwell, MA.

Petersen, E.R. (1971a). Bulk Service Queues: With Applications to Train Assembly Times. Working Paper 71-2, CIGGT, Queen's University, Kingston, Canada.

Petersen, E.R. (1971b). Queues with Random Batch Size with Applications to Railroad Modeling. Working Paper 71-77, CIGGT, Queen's University, Kingston, Canada.

Petersen, E.R. (1974). Over-the-Road Transit Time for a Single Track Railway. *Transportation Science*, 8(1):65–74.

Petersen, E.R. (1975a). A Primal-Dual Traffic Assignment Algorithm. *Management Science*, 22:87–95.

Petersen, E.R. (1975b). Interference Delays on a Partially Double-Tracked Railway with Intermediate Signalling. In *Proceedings of the Sixteenth Annual Meeting*. Transportation Research Forum.

Petersen, E.R. (1977a). Capacity of a Single Track Rail Line. Working Paper 77–38, CIGGT, Queen's University, Kingston, Canada.

Petersen, E.R. (1977b). Railyard Modeling: Part I. Prediction of Put-through Time. *Transportation Science*, 11(1):37–49.

Petersen, E.R. (1977c). Railyard Modeling: Part II. The Effect of Yard Facilities on Congestion. *Transportation Science*, 11(1):50–59.

Petersen, E.R. (1984). Rail Analysis Interactive Language (RAIL): A Decision Support System. In Florian, M., editor, *Transportation Planning Models*, pages 363–380. North-Holland, Amsterdam.

Petersen, E.R. and Fullerton, H.V. (eds) (1975). The Railcar Network Model. Working Paper 75-11, CIGGT, Queen's University, Kingston, Canada.

Petersen, E.R. and Taylor, A.J. (1993). A Structured Model for Rail Line Simulation and Optimization. *Transportation Science*, 16(2):192–206.

Pirkul, H. and Jayaraman, V. (1996). Production, Transportation, and Distribution Planning in a Multi-Commodity Tri-Echelon System. *Transportation Science*, 30(4):291–302.

Pirkul, H. and Jayaraman, V. (1998). A Multi-Commodity, Multi-Plant, Capacitated Facility Location Problem: Formulation and Efficient Heuristic Solution. *Computers & Operations Research*, 25(10):869–878.

Powell, W.B. (1981). *Stochastic Delays in Transportation Terminals: New Results in the Theory and Application of Bulk Queues*. PhD thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Powell, W.B. (1986a). A Local Improvement Heuristic for the Design of Less-than-Truckload Motor Carrier Networks. *Transportation Science*, 20(4):246–357.

Powell, W.B. (1986b). Analysis of Vehicle Holding and Cancellation Strategies in Bulk Arrival, Bulk Service Queues. *Transportation Science*, 19:352–377.

Powell, W.B. (1986c). Approximate, Closed Form Moment Formulas for Bulk Arrival, Bulk Service Queues. *Transportation Science*, 20(1):13–23.

Powell, W.B. (1986d). Iterative Algorithms for Bulk Arrival, Bulk Service Queues with Poisson and non-Poisson Arrivals. *Transportation Science*, 20(2):65–79.

Powell, W.B. (1987). An Operational Planning Model for the Dynamic Vehicle Allocation Problem with Uncertain Demands. *Transportation Research B: Methodological*, 21B:217–232.

Powell, W.B. (1988). A Comparative Review of Alternative Algorithms for the Dynamic Vehicle Allocation Problem. In B.L. Golden and A.A. Assad, editor, *Vehicle Routing: Methods and Studies*, pages 249–292. North-Holland, Amsterdam.

Powell, W.B. (1996a). A Stochastic Formulation of the Dynamic Assignment Problem, with an Application to Truckload Motor Carriers. *Transportation Science*, 30(3):195–219.

Powell, W.B. (1996b). Toward a Unified Modeling Framework for Real-Time Logistics Control. *Military Journal of Operations Research*, I(4):69–79.

Powell, W.B. (1998). On Languages for Dynamic Resource Scheduling Problems. In T.G. Crainic and G. Laporte, editor, *Fleet Management and Logistics*, pages 127–157. Kluwer Academic Publishers, Norwell, MA.

Powell, W.B. and Carvalho, T.A. (1998a). Dynamic Control of Logistics Queueing Networks for Large-Scale Fleet Management. *Transportation Science*, 32(2):90–109.

Powell, W.B. and Carvalho, T.A. (1998b). Real-Time Optimization of Containers and Flatcars for Intermodal Operations. *Transportation Science*, 32(2):110–126.

Powell, W.B., Carvalho, T.A., Godfrey, G.A., and Simaõ, H.P. (1995a). Dynamic Fleet Management as a Logistics Queueing Network. *Annals of Operations Research*, 61:165–188.

Powell, W.B. and Cheung, R.K.-M. (1994a). A Network Recourse Decomposition Method for Dynamic Networks with Random Arc Capacities. *Networks*, 24:369–384.

Powell, W.B. and Cheung, R.K.-M. (1994b). Stochastic Programs over Trees with Random Arc Capacities. *Networks*, 24:161–175.

Powell, W.B. and Cheung, R.K.-M. (2000). SHAPE - A Stochastic Hybrid Approximation Procedure for Two-Stage Stochastic Programs. *Operations Research*, 48(1):73–79.

Powell, W.B. and Frantzeskakis, L. (1994). Restricted Recourse Strategies for Dynamic Networks with Random Arc Capacities. *Transportation Science*, 28(1):3–23.

Powell, W.B. and Humblet, P. (1986). Queue Length and Waiting Time Transforms for Bulk Arrival, Bulk Service Queues with a General Control Strategy. *Operations Research*, 34:267–275.

Powell, W.B., Jaillet, P., and Odoni, A. (1995b). Stochastic and Dynamic Networks and Routing. In Ball, M., Magnanti, T.L., Monma, C.L., and Nemhauser, G.L., editors, *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*, pages 141–295. North-Holland, Amsterdam.

Powell, W.B. and Koskosidis, Y.A. (1992). Shipment Routing Algorithms with Tree Constraints. *Transportation Science*, 26(3):230–245.

Powell, W.B. and Shapiro, J.A. (2001). A Representational Paradigm for Stochastic, Dynamic Resource Transformation Problems. *Annals of Operations Research*.

Powell, W.B. and Sheffi, Y. (1983). The Load-Planning Problem of Motor Carriers: Problem Description and a Proposed Solution Approach. *Transportation Research A: Policy and Practice*, 17(6):471–480.

Powell, W.B. and Sheffi, Y. (1986). Interactive Optimization for Motor Carrier Load Planning. *Journal of Business Logistics*, 7(2):64–90.

Powell, W.B. and Sheffi, Y. (1989). Design and Implementation of an Interactive Optimization System for the Network Design in the Motor Carrier Industry. *Operations Research*, 37(1):12–29.

Powell, W.B., Sheffi, Y., Nickerson, K.S., Butterbaugh, K., and Atherton, S. (1992). Maximizing Profits for North American Van Lines' Truckload Division: A New Framework for Pricing and Operations. *Interfaces*, 18(1):21–41.

Raghavan, S. and Magnanti, T.L. (1997). Network Connectivity. In Dell'Amico, M., Maffioli, F., and Martello, S., editors, *Annotated Bibliographies in Combinatorial Optimization*, pages 335–354. John Wiley & Sons, New York, NY.

Regan, A.C. (1997). *Real-Time Information for Improved Efficiency of Commercial Vehicle Operations*. PhD thesis, University of Texas at Austin, Austin TX, U.S.A.

Regan, A.C., Mahmassani, H.S., and Jaillet, P. (1995). Improving the Efficiency of Commercial Vehicle Operations Using Real-Time Information: Potential Uses and Assignment Strategies. *Transportation Research Record*, 1493:188–198.

Regan, A.C., Mahmassani, H.S., and Jaillet, P. (1996a). Dynamic Decision Making for Commercial Vehicle Operations Using Real-Time Information. *Transportation Research Record*, 1537:91–97.

Regan, A.C., Mahmassani, H.S., and Jaillet, P. (1996b). Dynamic Dispatching Strategies under Real-Time Information for Carrier Fleet Management. In Lesort, J.B., editor, *Transportation and Traffic Theory*, pages 737–756. Pergamon.

Regan, A.C., Mahmassani, H.S., and Jaillet, P. (1998). Evaluation of Dynamic Fleet Management Systems: A Simulation Framework. *Transportation Research Record*, 1645:176–184.

Roy, J. (1984). *Un modèle de planification globale pour le transport routier des marchandises.* PhD thesis, Ecole des Hautes Etudes Commerciales, Université de Montréal, Montréal, Canada.

Roy, J. and Crainic, T.G. (1992). Improving Intercity Freight Routing with a Tactical Planning Model. *Interfaces*, 22(3):31–44.

Roy, J. and Delorme, L. (1989). NETPLAN: A Network Optimization Model for Tactical Planning in the Less-than-Truckload Motor-Carrier Industry. *INFOR*, 27(1):22–35.

Salkin, H.M. and Mathur, K. (1989). *Foundations of Integer Programming.* North-Holland, Amsterdam.

Séguin, R., Potvin, J.YJ., Gendreau, M., Crainic, T.G., and Marcotte, P. (1997). Real-time Decision Problems: An Operational Research Perspective. *Journal of the Operational Research Society*, 48:162–174.

Shan, Y. (1985). *A Dynamic Multicommodity Network Flow Model for Real-Time Optimal Rail Freight Car Management.* PhD thesis, Princeton University, Princeton, NJ, U.S.A.

Shilling, D.A. and Jayaraman, V. and Barkhi, R. (1993). A Review of Covering Problems in Facility Location. *Location Science*, 1:25–55.

Skorin-Kapov, D. and Skorin-Kapov, J. (1994). On Tabu Search for the Location of Interacting Hub Facilities. *European Journal of Operational Research*, 73:502–509.

Toth, P. and Vigo, D., editors (2002). *The Vehicle Routing Problem*, volume 9 of *SIAM Monographs on Discrete Mathematics and Applications.* SIAM.

Turnquist, M.A. (1994). Economies of Scale and Network Optimization for Empty Car Distribution. Technical report, Cornell University.

Turnquist, M.A. and Daskin, M.S. (1982). Queuing Models of Classification and Connection Delay in Railyards. *Transportation Science*, 16(2):207–230.

Turnquist, M.A. and Markowicz, B.P. (1989). An Interactive Microcomputer-Based Planning Model for Railroad Car Distribution. Technical report, Cornell University.

Van Roy, T.J. (1986). A Cross Decomposition Algorithm for Capacitated Facility Location. *Operations Research*, 34:145–163.

White, W. (1972). Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers. *Networks*, 2(3):211–236.

White, W.W. (1968). A Program for Empty Freight Car Allocation. Technical Report 360D.29.002, IBM Contributed Program Library, IBM Corporation, Program Information Department, Hawthorne, N.Y.

White, W.W. and Bomberault, A.M. (1969). A Network Algorithm for Empty Freight Car Allocation. *IBM Systems Journal*, 8(2):147–171.

Winston, C. (1983). The Demand for Freight Transportation: Models and Applications. *Transportation Research A: Policy and Practice*, 17A:419–427.

Yang, J., Jaillet, P., and Mahmassani, H.S. (2002). Study of a Real-Time Multi-Vehicle Truckload Pick-Up and Delivery Problem. *Transportation Science*. forthcoming.

Yang, J., Mahmassani, H.S., and Jaillet, P. (1999). On-Line Algorithms for Truck Fleet Assignment and Scheduling under Real-Time Information. *Transportation Research Record*, 1667:107–113.